Hindawi
Scientific Programming
Volume 2022, Article ID 6177978, 14 pages
https://doi.org/10.1155/2022/6177978

Research Article

Energy Management Model of Intelligent Park Based on Improved Depth Deterministic Gradient Strategy Algorithm

Chen Xu[1] and Xianwei Jiang [2]

[1] *Department of Mechanical and Electrical Engineering, Suzhou Vocational and Technical College, Suzhou, Anhui Province 234000, China*
[2] *Anhui Province Key Laboratory of Simulation and Design for Electronic Information System, Hefei Normal University, Hefei, Anhui Province 230601, China*

Correspondence should be addressed to Xianwei Jiang; xianweij@hfnu.edu.cn

Received 9 November 2021; Revised 17 December 2021; Accepted 18 December 2021; Published 14 February 2022

Academic Editor: Sikandar Ali

The traditional distribution automation system, demand side management system, and distributed generation access and control system, respectively, solve the problems of regional distribution network power supply, customer power consumption, and new energy utilization to varying degrees. Intelligent Park evolved from the concept of intelligent park of State Grid Corporation, Modern enterprises, or residential areas that can make full use of modern communication, computer, automation, and other technologies to implement the power supply demand of the park. For the modern intelligent park with user-side temperature control load and demand response load access, an intelligent park energy management and optimal scheduling method based on deep reinforcement learning (DRL) algorithm is proposed. Through the interaction between the agent and the energy environment of the intelligent park, the control strategy is adaptively learned. This method can realize the continuous action control of energy management in intelligent park and can realize the optimal scheduling decision of intelligent park in various scenarios. Firstly, based on the characteristics and types of intelligent park aggregation unit, the environment model of intelligent park energy management system interacting with agent is established. Secondly, the basic principle of deep deterministic policy gradient deep algorithm is introduced, and on this basis, the key links of deep reinforcement learning algorithm, such as action space, state space, reward mechanism, neural network structure, and learning process, are designed. Finally, the effectiveness of the proposed algorithm is verified by an example of a city intelligent park.

1. Introduction

As a typical Internet of Things [1–6] application, intelligent park refers to a regional autonomous system composed of distributed photovoltaic and energy storage units on the user side and various flexibly adjusted loads, which coordinates with the distribution network or independently supplies power to the loads [7–9]. In recent years, intelligent parks have shown great potential in grid connection and consumption of distributed new energy, aggregation, and regulation of controllable load on the user side and have become one of the main strategies for developing distributed energy and improving energy utilization rate on the user side in many developed countries [10]. Different from the traditional microgrid concept, smart parks add flexible and controllable resources such as air conditioners and electric vehicles to their aggregation units, which aggravates the uncertainty and complexity of commercial transactions and power logistics in smart grids to a certain extent [11]. Therefore, how to understand personal electricity consumption behavior and its impact from a large number of high-dimensional converged data on the user side and fully mobilize the flexibility of user-side resources has become an urgent problem to be solved in microgrid energy management [12, 13].

At present, for the energy management and optimal scheduling of intelligent parks, from the perspective of algorithm, most of them focus on optimization algorithm or heuristic intelligent algorithm. Sun et al. in [14] established an optimal dispatching model of intelligent park considering photovoltaic, ice storage air conditioning load, and interruptible load on the user side and established a mixed integer programming model of intelligent park. On this basis, the power flow constraint of distribution network is further considered in [15], and the energy management model of microgrid is decomposed into a unit commitment problem and an optimal power flow problem, which avoids solving the mixed integer nonlinear programming problem directly. Zhang et al. in [16] further studied the optimal scheduling problem of microgrid under uncertain photovoltaic output and proposed a two-stage robust optimization model of microgrid based on C-CG algorithm. Lei et al. in [17] studied multi-microgrid complementary scheduling and energy sharing on the basis of single microgrid operation optimization, established a master-slave game model between multi-microgrid and distribution network, and used Kriging metamodel fitting with less computation to replace the energy management model of lower microgrid. On this basis, Zhao and Tao et al. in [18, 19] further study the distributed optimization of multi-microgrid and realize distributed computing of multi-microgrid based on alternating direction multiplier method, which protects the privacy of each microgrid. Askarzadeh in [20] uses a genetic algorithm based on memory mechanism to directly solve the minimum operating cost problem of microgrid. In [21], hybrid particle swarm optimization algorithm is used to solve the multiobjective optimization problem considering both distribution network security and microgrid operation economy.

The research methods of optimizing the operation of intelligent parks in the above documents can be summarized as centralized and distributed solutions. Specifically, when using these traditional methods for optimization, all possible solutions must be calculated completely or partially, and the best solution must be selected. However, this kind of calculation leads to the low computational efficiency of the traditional optimization process, and when the optimization model is nonlinear, it is easy to fall into the local optimal solution. Relying on the background of big data, deep reinforcement learning (DRL) algorithm provides a new way to solve this problem. As an artificial intelligence scheme, DRL algorithm has been widely used in the fields of optics, communication, geography, and power system scheduling. She et al. in [22] applied unsupervised deep learning and deep reinforcement learning to 6G mobile communication, which verifies the effectiveness of different algorithms. The deep Q-learning (DQN) algorithm is applied to cooperative communication technology [23], and DQN is trained according to interrupt probability and mutual information, and the optimal relay is selected from multiple relay nodes without network model or prior data. Cao et al. in [24] used the deep reinforcement learning algorithm to optimize the arbitrage strategy of energy storage system and forecasted the load data by combining convolution neural network and long-term memory. Li et al. and Qiu et al. in [25, 26] further applied deep reinforcement learning to the charge and discharge optimization of electric vehicles.

As a model-free optimization algorithm, DRL has been widely used in the fields of demand response optimization, power market decision optimization, and power system optimal dispatching. Wang et al. in [27] designed a model-free deep reinforcement learning method with double DQN structure to optimize the demand response management under time-of-use electricity price and variable power consumption mode; the actor-critical algorithm is used to study the demand response regulation of load aggregation quotient, and the effectiveness of the proposed algorithm is proved under the condition of considering various uncertain factors [28]; Ye et al. in [29] studied the optimization of electricity market clearing and used deep deterministic policy gradient (DDPG) algorithm to solve the bilevel model. Liu et al. in [30] used A3C algorithm to realize residents' load participating in demand side response and made use of CPU multithreading function to execute multiple actions at the same time through multiagent. It can be seen that DRL can learn from the continuous transformation of historical data. This powerful machine learning model, combined with deep neural network, can quickly extract, control, and optimize the terminal aggregation units of intelligent park, and can deal with the high uncertainty under various complex modes. The DRL algorithm has been applied to various fields of power system from different angles in the literature mentioned above, but in general, there are still some shortcomings as follows: (1) Although the existing DRL algorithm can improve the convergence by designing the target network, experience playback, and other mechanisms, its training is slow and converges. Difficulties, instability, and other problems are still widespread, and it is difficult to adjust parameters; (2) although some researches have applied DRL algorithm to the cooptimal scheduling of the park, the park model is simple and often ignores the user-side demand response resources, such as temperature-controlled load, transferable/interruptible load, etc., so it is difficult to reflect the benign interaction of source-load. Based on the above considerations, this paper improves the existing DDPG algorithm from the perspectives of random exploration strategy, environment awareness and learning ratio, and increasing experience playback pool and applies the improved DDPG algorithm to the energy management and optimal scheduling decision of modern intelligent parks. Finally, an example of an actual intelligent park in a city is given to verify the effectiveness of the proposed algorithm in different scenarios.

To sum up, the main contributions of this paper are as follows:

(1) Aggregate modeling of user-side multishape flexible regulation resources in intelligent park, construct user-side temperature control load model and price demand response model, and take user-side regulation into account in the energy management model of intelligent park.

(2) An energy management model based on improved depth deterministic gradient strategy is proposed in this paper. Through such three measures as the random exploration strategy, the adjustment of environment perception and learning proportion,

and the increase of high access experience pool, the convergence characteristics and computing ability of the algorithm are improved significantly.

(3) The effectiveness of the proposed algorithm is verified by an example of an actual intelligent park in a city, and the improvement of the overall economic benefit of the park by controlling user-side resources is verified by a comparative example.

2. Intelligent Park Structure and Scheduling Framework

The microgrid studied in this paper is a community microgrid with independent supply and demand infrastructure. The microgrid is managed by an energy aggregator or a public utility company. There are wind turbines in the microgrid, which are connected to the main power grid and can supply internal load or sell electricity to the superior power grid according to the market electricity price. The overall architecture of the intelligent park is shown in Figure 1, its architecture can be analyzed from three angles: physical layer, information layer, and control layer. The physical layer includes a wind turbine, an electric energy storage system, a set of temperature-controlled loads, and a set of price-based demand response loads. The information layer is mainly composed of smart meters and communication systems, which can realize real-time monitoring of power generation and consumption data and bidirectional transmission of information of various components. The control layer is mainly an energy management system (EMS) that sends control signals to the controllable components of the microgrid through the relevant infrastructure. As can be seen from Figure 1, the main control variables include (1) switch control quantity of temperature control load; (2) charge/discharge control of electric energy storage system; (3) interactive power control with superior power grid. Therefore, when modeling energy management in this paper, a multiagent control system can be established according to these control variables. These control units will be modeled one by one below.

2.1. Modeling of Electric Energy Storage System. Considering the economic and technical feasibility, this paper adopts a community shared energy storage system, which can meet the power demand of intelligent park for at least two hours. The real-time stored power status of the energy storage system is expressed as follows:

$$B_t = B_{t-1} + \eta_c C_t - \frac{D_t}{\eta_d}, \tag{1}$$

where ${}^{t}_{B} \in [0, {}^{\max}_{B}]$ is the electric energy stored by the electric energy storage system at time t, ${}^{\max}_{B}$ is the maximum capacity of the electric energy storage system, ${}^{c}_{\eta}$ and ${}^{d}_{\eta}$, respectively, represent the charging efficiency and discharging efficiency of the energy storage system, ${}^{t}_{C} \in [0, {}^{\max}_{C}]$ and ${}^{t}_{D} \in [0, {}^{\max}_{D}]$, respectively, represent the charging power and discharging power of ESS at time t, and ${}^{\max}_{C}$ and ${}^{\max}_{D}$, respectively, represent the maximum charging power and discharging power of ESS. The storage state of the ESS system can be expressed by SOC values as

$$BSC_t = \frac{B_t}{{}^{\max}_{B}}. \tag{2}$$

2.2. Temperature Control Load Modeling. The temperature control load in the intelligent park refers to a large number of air conditioners, water heaters, heat pumps, or refrigerators, which can form flexible adjustment resources with a certain controllable capacity through aggregators and can participate in the resource scheduling of the intelligent park as a whole under the coordinated control of agents. Its basic control structure is shown in Figure 2.

In order to ensure the comfort level of users, a feedback controller is equipped on each TCL to maintain the temperature of related equipment within an acceptable range. The feedback controller can receive the switch action information of the equipment from the aggregator and verify whether the temperature constraint can be met under the action. The action control strategy of the feedback controller is as follows:

$$u^i_{t+1} = \begin{cases} 0 & if\ T^i_t > T^i_{\max} \\ u^i_t & if\ T^i_{\max} > T^i_t > T^i_{\min} \\ 1 & if\ T^i_t < T^i_{\min} \end{cases}, \tag{3}$$

where u^i_t is the decision variable of the feedback controller, indicating the off-state of the equipment; if it is in the open state, it is set to 1; otherwise it is set to 0; T^i_t represents the temperature of the i-th temperature control load at time t, and $T^i_{\min}$ and $T^i_{\max}$ are the minimum and maximum values of temperature, respectively. The temperature dynamic change process of temperature control load satisfies the second law of thermodynamics [31], which is specifically expressed as follows:

$$\begin{aligned} T^i_{t+1} &= \frac{1}{C^i_a}\left(T^0_t - T^i_t\right) + \frac{1}{C^i_m}\left(T^i_{m,t} - T^i_t\right) + L^i_{tcl}u^i_t + q^i. \\ T^i_{m,t+1} &= \frac{1}{C^i_m}\left(T^i_t - T^i_{m,t}\right), \end{aligned} \tag{4}$$

where T^i_t is the measured indoor air temperature, $T^i_{m,t}$ is the unobservable building temperature, T^0_t is the outdoor temperature, C^i_a and C^i_a are the specific heat capacity of air and buildings, ${}^{i}_{q}$ is the internal heat of buildings, and L^i_{tcl} is the rated power of temperature control load. Similarly, the adjustable range of temperature control load can also be expressed in the form of state of charge, specifically as follows:

$$SoC^i_t = \frac{T^i_t - T^i_{\min}}{T^i_{\max} - T^i_{\min}}. \tag{5}$$

2.3. Price-Based Demand Response Load Modeling. Price-based demand response load refers to some interruptible/transferable loads in intelligent parks, which can participate in the energy scheduling of microgrid on the

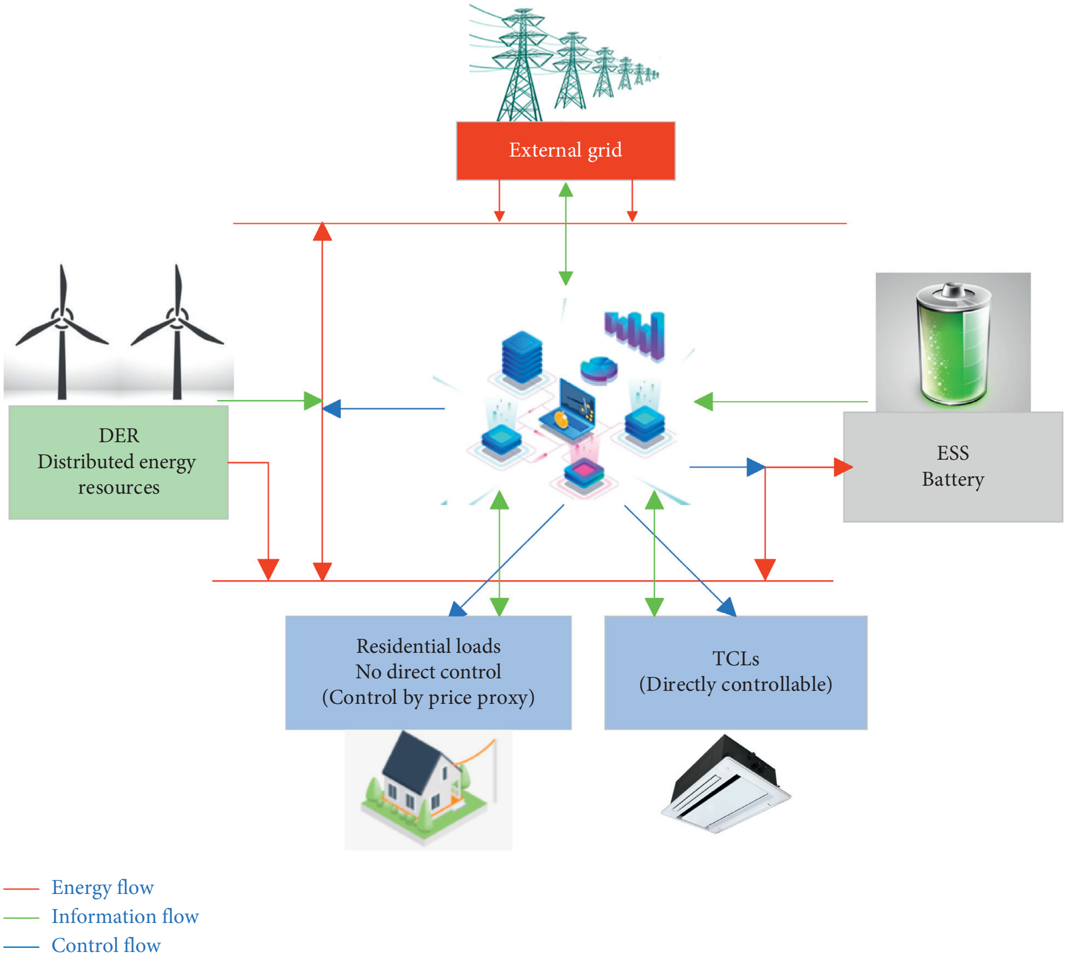

Figure 1: Overall structure of community micronetwork.

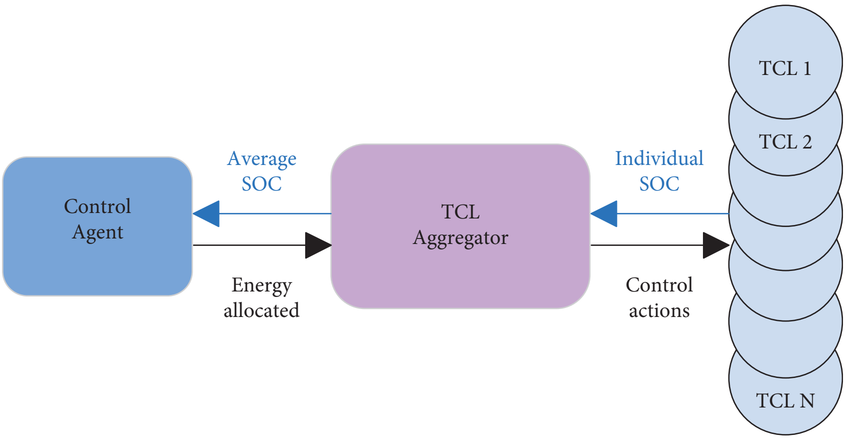

Figure 2: TCL aggregation control chart.

premise of meeting the minimum demand of users according to the electricity price. Based on sensitivity coefficient $\beta \in [0, 1]$ and recovery coefficient $^{i}_{\lambda}$, this paper describes the characteristics of demand response load of different households. The so-called sensitivity coefficient refers to the load percentage that response load I can be reduced/increased under the condition of given electricity price increase/decrease; the so-called recovery coefficient refers to how many times the reduced load can be supplemented. The specific load size of price demand response load I at time t is

$$\begin{aligned} L_t^i &= L_{b,t} - SL_t^i + PB_t^i. \\ SL_t^i &= L_{b,t} \cdot \beta_i \cdot \delta_t. \end{aligned} \tag{6}$$

In the formula, $L_{b,t}$ tables the basic load size, SL_t^i represents the actual load transferred at time t, the specific expression is shown in (6), $\delta_t = \{-2, -1, 0, 1, 2\}$ represents the load electricity price level at time t, $\delta_t > 0$ at that time, and the electricity price is higher at that time. EMS reduces the operating cost by transferring the load, so $SL_t^i > 0$ at this time; similarly, $\delta_t < 0$ at that time and $SL_t^i < 0$ at this time. PB_t^i represents the compensation amount of the previously transferred load at time t, and its specific calculation method is as follows:

$$PB_t^i = \sum_{j=0}^{t-1} w_{i,j} SL_j^i, \tag{7}$$

where $w_{i,j} \in \{0, 1\}$ represents whether load transfer occurs at time, and this variable is a Boolean variable, which is mainly affected by electricity price level and recovery coefficient. Specifically, when the transfer duration is closer to the maximum recovery coefficient, the greater probability of $w_{i,j} = 1$, and the actual probability of $w_{i,j} = 1$ at time t can be carried out in the following way:

$$\begin{gathered} P\left(w_{i,j} = 1\right) = clip\left(\frac{-\delta_t \cdot sign\left(SL_j^t\right)}{2} + \frac{(t-j)}{\lambda_i}, 0, 1\right). \\ clip(X, a, b) = \begin{cases} a & if\, X < a \\ X & if\, a \le X \le b. \\ \mathrm{b} & if\, X > b \end{cases} \end{gathered} \tag{8}$$

The intelligent park is also connected with the superior power grid through tie lines to balance the internal supply and demand relationship in real time. That is to say, when the internal supply and demand relationship cannot be balanced, the superior power grid can be used as the standby dispatching resource of the microgrid, and its transaction price adopts the Finnish electricity market price provided by [32]. The agent of the main power grid needs to share the purchase/sale price information with EMS in real time, which is recorded as $(\lambda_t^u, \lambda_t^d)$. The agent control flow is shown in Figure 3, and the specific steps are as follows:

(1) According to the electricity price level, determine the basic electricity consumption strategy of demand response load and temperature control load.

(2) Determine the relationship between energy supply and demand according to the demand response, the actual electricity consumption of temperature-controlled load, and the output of wind turbines.

(3) According to the relationship between energy supply and demand, priority is given to scheduling energy storage to reduce power imbalance, and if the supply is in short supply, energy storage will discharge. If the supply exceeds the demand, the excess energy will be used for energy storage and charging.

(4) If the energy balance cannot be realized after (3), the purchase/sale of electricity from the superior power grid shall be considered, and the purchase/sale value shall be recorded as (P_t^{buy}, P_t^{sell}) until the energy balance is realized.

EMS should fully consider the price response characteristics of various resources and determine the load electricity price level δ_t at all times in the intelligent park, and in order to prevent the actual owner of the park from maliciously raising the electricity price, it is stipulated that the quotation of EMS should fluctuate around an intermediate value, and at the same time, the daily average price level should be close to the retail electricity price provided by power retailers. See the following formula for the specific form of EMS quotation:

$$P_t \in \left(P_{\text{market}} + \delta_t \cdot cst\right)_{\delta_t \in \{-2,-1,0,1,2\}}. \tag{9}$$

In the formula, in order to reflect a constant of price change speed, ${}^{market}_{P}$ is the retail price. In the actual operation process, EMS continuously records the price level for each short time and accumulates it. When the price level exceeds a certain threshold, the retail price is used as the actual transaction price. The final price level calculation formula is as follows:

$$\delta_{t,eff} = \begin{cases} \delta_t & if \sum_{j=0}^{t} \delta_t \le \text{threshold} \\ 0 & if \sum_{j=0}^{t} \delta_t \le \text{threshold} \end{cases}, \tag{10}$$

where threshold is the preset threshold.

3. Reinforcement Learning Framework for Energy Management in Intelligent Parks

The energy management problem of intelligent park is essentially a sequential decision-making problem, and its mathematical essence is stochastic dynamic programming. At present, most of the common methods to solve this kind of problems are based on fine model, which depend on various numerical solutions or heuristic intelligent acid and often require the objective function to be derivative or differentiable. The reinforcement learning algorithm adopted in this paper is a data-driven model-free algorithm. Adaptive learning is carried out by means of "trial and error" of agents. Through continuous interaction with the environment, constantly obtain the environmental state and take corresponding actions to change the environmental state according to certain updating strategies. In this process, the agent will get certain rewards or punishments, and the agent will use the reward value/punishment value as the updating guidance of model parameters, in order to get the maximum cumulative rewards in the process of continuous learning.

3.1. Markov Decision Process. The above-mentioned process of perception-action-evaluation-learning is also called Markov decision process (MDP). The four basic elements of

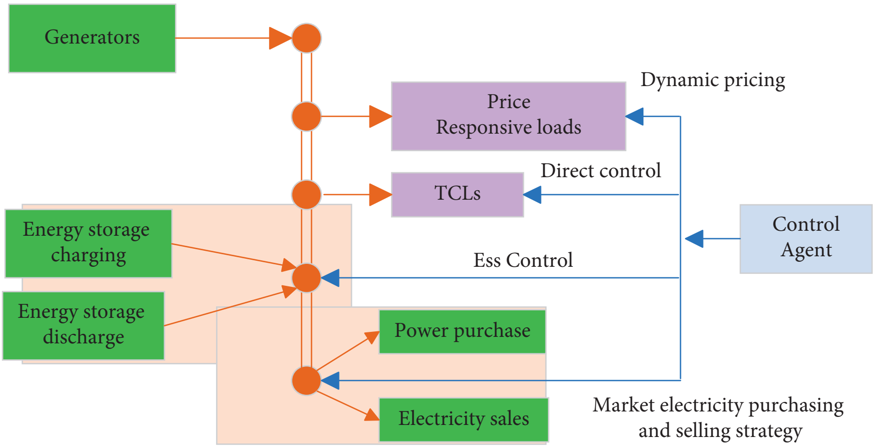

Figure 3: Basic control structure of EMS.

MDP are state space, action space, transition function, and reward function. Among them, transition function describes the probability of agent changing from state $s \in S$ to $s' \in S$ under a given action:

$$\begin{gathered} T\colon SAS \longrightarrow [0,1] \\ T(s,a,s') = p(s_{t+1} = s' | s_t = s, a_t = a). \end{gathered} \tag{11}$$

At every moment, the agent selects the corresponding action in the possible action set A according to the state s_t and the random strategy π (the strategy π is a function of mapping the state s_t to the action a). In return, the agent receives the understood reward value R_t of the process and updates it to the next state variable s_{t+1}. The ultimate goal of the agent under the policy π is to maximize the accumulated discount reward value, which is expressed as follows:

$$\rho_t = \sum\nolimits_{k=t}^{D-1} \gamma^{k-t} R_k. \tag{12}$$

In the formula, $\gamma \in [0,1]$ is the discount factor, which directly determines the influence degree of rewards in subsequent steps on the current period. According to the strategy π, the expected value of rewards under the current strategy can be obtained as follows:

$$V^{\pi}(s_t) = \overset{\pi}{E}(R_t + \gamma V^{\pi}(S_{t+1})), \forall t < D - 1. \tag{13}$$

The Q value function is the action value function, which indicates the expected return after selecting the action a_t:

$$Q^{\pi}(s_t, a_t) = R_t + \gamma V^{\pi}(s_{t+1}). \tag{14}$$

Starting from the state s_0, the agent begins to search for the best strategy step by step, which maximizes the action value function:

$$\pi_* = \arg\max_{\pi} Q^{\pi}(s,a), \tag{15}$$

$$Q^{\pi=*}(s,a) = Q^{\pi^*}(s,a). \tag{16}$$

From equations (14)–(19), it can be seen that the model-free reinforcement learning method does not need to know the specific expression of the state transition probability function T but only needs to train according to the transition process of each interaction with the environment, so that the expected value of the total reward can reach the highest. The energy management and optimal dispatching method of power grid is transformed into Markov decision model. Then the rule-based dynamic energy management method is selected as the basic strategy, which can reduce the overall optimization cost, so as to avoid solving the optimal value function of all states. Then the adaptive learning control strategy is proposed, which can realize the continuous action control of energy management in intelligent park and the optimal dispatching decision in various scenarios.

3.2. Reinforcement Learning Framework for Energy Management in Intelligent Parks.

In this paper, the intelligent park system is the agent's environment. The agent optimizes the scheduling by adjusting the output and power generation of related aggregation units in the system. In time T, the environment provides the observed system state $s_t \in S$ to the agent, and the agent generates action DD based on strategy $s_t \in S$ and intelligent park system state $a_t \in A$.

3.2.1. State Space.

The state space is composed of the reference information of the decision-making of the agent at each moment, which mainly includes controllable state variable S^C and external state variable S^X. Among them, the controllable state variables are mainly environmental variables that can be directly/indirectly controlled by agents, such as the average SOC value of temperature-controlled load, the maximum SOC value of ESS, and the load electricity price level δ_t. External state variables are variables beyond the control of agents, such as temperature T_t and wind turbine power generation and higher-level market purchase and sale price $^{t}_{G}$ and P_t^u, P_t^d. Therefore, the state space can be described as

$$\begin{gathered} s_t \in S = S^C \mathrm{x}\, S^X \\ s_t = \left[SoC_t, BSC_t, G_t, \lambda_t^u, \lambda_t^d, L_{b,t}, t\right]. \end{gathered} \tag{17}$$

3.2.2. Action Space. The action space of the agent is mainly composed of four parts, including temperature control load action space A_{tcl}, load price action space A_P, energy storage charge and discharge action space A_D, and market purchase and sale electricity action space A_E, which are specifically expressed as follows:

$$\begin{aligned} a_t \in A &= A_{tcl} \text{x}\, A_p \text{x}\, A_D \text{x}\, A_E \\ a_t &= \left[P_t, u_{t+1}^i, C_t, D_t, P_t^{buy}, P_t^{sell}\right]. \end{aligned} \tag{18}$$

3.2.3. Reward Function. The reward function is set to maximize the operational benefits of the intelligent park. The calculation method is to subtract the power generation cost from the income from selling electricity to the external power grid and the cost of purchasing electricity from the external power grid. The expression of the reward function is as follows:

$$\begin{aligned} R_t &= S_t - C_t. \\ S_t &= P_t \sum_{loads} L_t^i + C_{gen} \sum_{tcls} L_{tcl}^i u_t^i + \lambda_t^d P_t^{sell}. \\ \text{Costs}_t &= C_{\text{gen}} G_t + \left(\lambda_t^u + C_{\text{trimp}}\right) P_t^{buy} + C_{tr\,\text{exp}} P_t^{sell}, \end{aligned} \tag{19}$$

where λ_t^u, λ_t^d are lower standby and upper standby electricity prices, which are the transaction electricity prices between microgrid and superior power grid. G_t is the power generation of wind turbines, C_{gen} is the power generation cost, and C_{trimp} and $\overset{tr\,\text{exp}}{C}$ are the energy transmission cost of purchasing electricity from or selling electricity to the power grid.

3.3. Energy Management Framework of Intelligent Park Based on DDPG Algorithm. The traditional reinforcement learning algorithm based on *Q*-learning performs well in dealing with small-scale discrete space problems. However, when dealing with practical problems with continuous state variables, with the increase of space dimensions, the discrete state variables will increase exponentially, which brings serious dimension disaster and cannot realize efficient learning of agents. In the energy management of intelligent park in this paper, firstly, the state variables such as wind turbine output, user load value, and market trading electricity are all continuous variables. In addition, the market transaction electricity quantity and energy storage charging and discharging power in the action space are also continuous variables, so the traditional reinforcement learning algorithm cannot effectively solve the energy management problem of intelligent park.

Based on the above considerations, this paper uses deep neural network (DNN) to approximate reinforcement learning, so as to solve the energy management problem of intelligent park with continuous state/action space [33]. The DDPG algorithm based on the actor-critical framework is adopted in this paper. It uses two independent networks to approximate the critical function (θ^Q) and the actor function (θ^Q), and each network has its own target networks $\theta^{Q\prime}$ and $\theta^{\pi^{\backprime}}$, for which the target Q value and target strategy are denoted as $Q\prime$ and π'.

3.3.1. Value Network Training. For value networks, the network parameters are usually optimized by minimizing loss function:

$$L\left(\theta^Q\right) = E\left(y_t - Q\left(s_t, a_t|\theta^Q\right)\right)^2, \tag{20}$$

where y_t is the target Q value; the calculation method is as follows:

$$y_t = r_t + \gamma Q'\left(s_{t+1}, \pi'\left(s_{t+1}|\theta^{\pi'}\right)|\theta^{Q'}\right). \tag{21}$$

The value network parameters are updated in the gradient direction, and $L(\theta^Q)$ is calculated as follows about θ^Q gradient:

$$\nabla_{\theta^Q} L\left(\theta^Q\right) = E\left(2\left(y_t - Q\left(s_t, a_t|\theta^Q\right)\right)\nabla_{\theta^Q} Q\left(s_t, a_t\right)\right). \tag{22}$$

According to the gradient rule, the update formula of the value network parameters is as follows:

$$\theta^Q \leftarrow \theta^Q - \mu_Q \nabla_{\theta^Q} L\left(\theta^Q\right). \tag{23}$$

3.3.2. Strategic Network Training. The training and updating of the strategy network are also based on gradient information. The gradient information of the strategy network is recorded as $\nabla_a Q(s_t, a_t|\theta^Q)$, and the sampling strategy gradient is calculated as follows:

$$\nabla_{\theta^\pi} \pi = \nabla_a Q\left(s, a|\theta^Q\right)|_{s=s_t, a=\pi(s_t)} \nabla_{\theta^\pi} \pi\left(s, |\theta^\pi\right)|_{s=s_t}. \tag{24}$$

According to the above deterministic policy gradient, the policy network parameters can be updated:

$$\theta^\pi \leftarrow \theta^\pi - \mu_\pi \nabla_{\theta^\pi} \pi, \tag{25}$$

where μ_π is the strategic network learning rate.

The target network parameters $\theta^{Q'}$ and $\theta^{\pi'}$ sampling soft update technology further improve the stability of the learning process:

$$\begin{aligned} \theta^{Q'} &\leftarrow \tau\theta^Q - (1-\tau)\theta^{Q'}. \\ \theta^{\pi'} &\leftarrow \tau\theta^\pi - (1-\tau)\theta^{\pi'}, \end{aligned} \tag{26}$$

where τ is the soft renewal coefficient.

3.4. Improved DDPG Algorithm

3.4.1. Improvement of Convergence Mechanism

(1) Random Exploration. In this paper, random noise is superimposed on the output actions to explore the environment; that is, in the process of training, the actions taken every time are improved as follows:

$$\widehat{a}_t = p a_t + (1-p) n, \tag{27}$$

where $\widehat{a}_t$ is the actual action after updating, and n is the noise that obeys the truncated normal distribution within the range of $[-1, 1]$; p is the proportion of neural network output action value a_t. The larger the value, the lower the randomness of exploration. In order to avoid any exploration at all, the upper limit of its value is 0.95.

(2). Adjust the Proportion of Environmental Perception and Learning. In the traditional DDPG algorithm, every time an agent interacts with the environment, it needs to learn the model parameters once. In fact, this frequent learning time greatly increases the training time. Without sufficient interaction and exploration of the environment, frequent learning can easily make the agent fall into the local optimal value. In this paper, we adjust this and set the agent to learn every 30 times when it interacts with the environment.

3.4.2. Adding a Playback Pool for High Access Experience. Due to the high proportion of superimposed exploration in the output actions in the early stage of model training, there are few transfer strategies with higher total return value, and the model training speed is slow, and the influence of random exploration is even higher than that of scheduling sequence itself. In order to improve the convergence speed of the algorithm, an experience playback pool for storing high-quality transfer process is added, and its admission condition is set to be higher than the average value of the total reward value in each training cycle, so as to accelerate the convergence speed of the previous algorithm.

In fact, the three measures basically do not change the complexity of the algorithm from the mathematical model, only by adjusting the relevant parameters, or some of the results have been trained to optimize, because the mathematical model does not increase the complexity of the algorithm.

In summary, the basic framework of solving energy management problems in intelligent parks based on improved DDPG algorithm is shown in Figure 4. Under this framework, the input is state variables, including market electricity price information, wind turbine output, and other state variables, and the output is action vectors, which mainly include action variables such as charging and discharging power of electric energy storage system, power consumed by TCL, demand response load, and market purchase and sales power.

(3). Example Test and Result Analysis. In this paper, a toolkit named Gym of Open AI is used to build the simulation environment of intelligent park. TensorFlow 1.1.4 Toolkit is used to train neural network, and the training data of wind power and market electricity price are taken from [34].

In terms of model parameter setting, the charging and discharging efficiency of ESS is 0.9, the maximum charging and discharging power is set to 250 kW, the capacity is set to 500kWh, the wind power generation cost is 32/MW, the upper standby price and the lower standby price are taken from [35], C_{trimp} and $\overset{tr\,\exp}{C}$ are 0.9/MW and 1.3/MW, respectively, and TCL parameters are taken from [36]. The total quantity is 100, the total period of a day is 24 h, and the retail electricity price is 5.48 euro/kW. All experimental results are average by multiple times.

The structure of DDPG network is shown in Appendix in Figures 5 and 6. Considering the dimensions of state space and action space, the number of hidden layer neurons of actor network is 300/100/100, and the number of hidden layer S_1, A_1, H_1, H_2 neurons of critic network is 200, 50, 100, and 100, respectively. ReLU function is used as the activation function of hidden layer. The learning rate of actor network and critic network is 0.001 and 0.0001, respectively, the learning rate of target network is 0.001, and the capacity of experience pool is 10^5.

3.5. Comparison of Convergence Speed and Learning Effect of Different Algorithms. In order to show the effectiveness of the improved DDPG algorithm proposed in this paper, the following basic comparative examples are set, as shown in Table 1.

As can be seen from the table, compared with scheme 1 and scheme 2, it can be seen that the reward value is increased by about 19.75% and the training time is reduced by nearly 2 hours after adopting the random exploration strategy proposed in this paper. In addition, comparing scheme 3 and scheme 2, set the ratio of environment awareness and agent learning to 30: After 1, not only has the reward value of the agent been greatly increased, but also the training time has been obviously reduced, which is mainly due to the improvement of environment perception and agent learning ratio, and the reduction of a large number of unnecessary learning times, so the convergence speed of the algorithm is significantly improved, and the agent is prevented from falling into local optimal solution. Finally, compared with scheme 4 and scheme 3, it can be seen that, after adding experience pool, the algorithm further improves the convergence speed and reduces the training time on the premise of ensuring a high reward value. The reward curves under scenarios 1, 2, 3, and 4 are shown in Figure 7.

Figure 7 shows the reward curves of the four different schemes in Table 1, with different scheme configurations resulting in different curve trends. Under the random exploration strategy, further refine the final reward value of agents under different perceptual learning ratios, as shown in Figure 8. As can be seen from the figure, although the perceptual learning ratio can greatly reduce the training time of the model, avoid the model falling into local optimum, but when the training is carried out to a certain extent, too high proportion of perceptual exploration will lose more key information. As a result, the overall reward value is reduced and the training time is increased. Under the four comparison schemes, when the perceptual learning ratio is 30 : 1, the training time and the final reward value are optimal. Therefore, in the process of adjusting the parameters of the time model, it is necessary to set the perceptual learning ratio more in line with the model training in combination with practical problems.

In order to demonstrate the effectiveness of the improved DDPG algorithm compared with the existing one, the experimental results of the existing algorithms compared with the DQN algorithm, the actor-critic algorithm, and the PPO algorithm under different hyperparameters are tested respectively, as shown in Table 2.

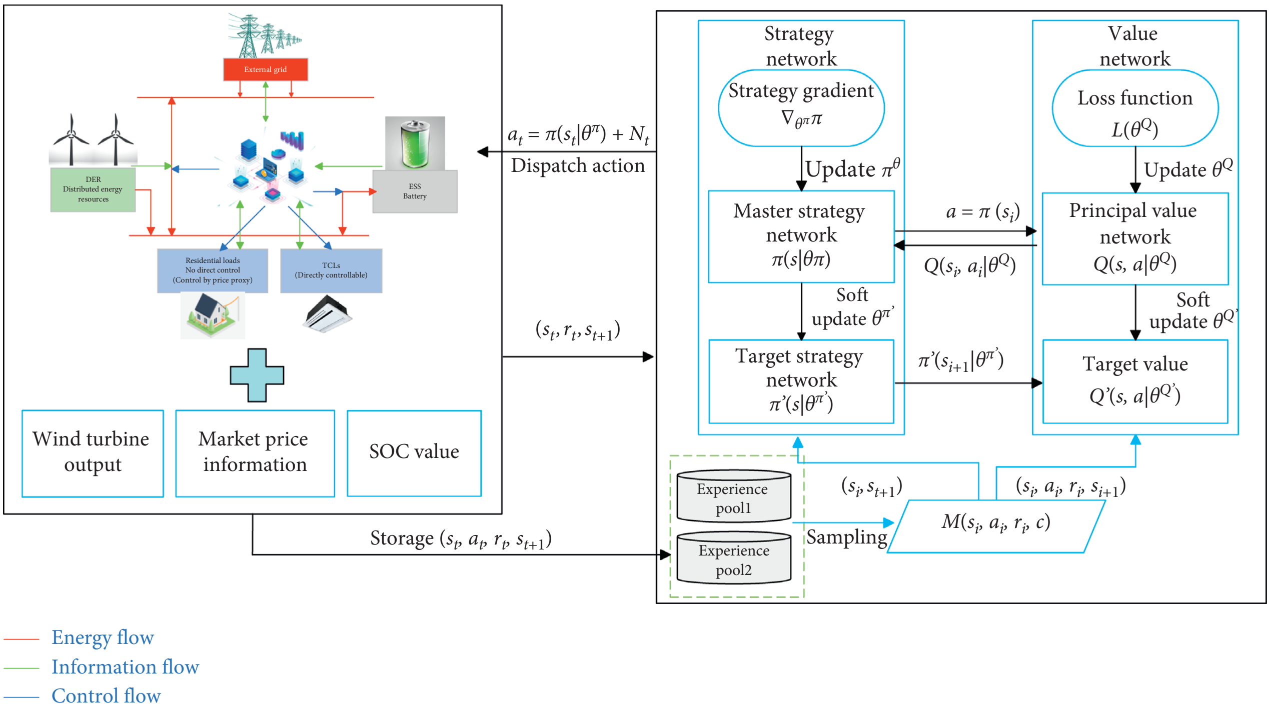

Figure 4: Structure diagram of community microgrid optimal scheduling model based on improved A3C algorithm.

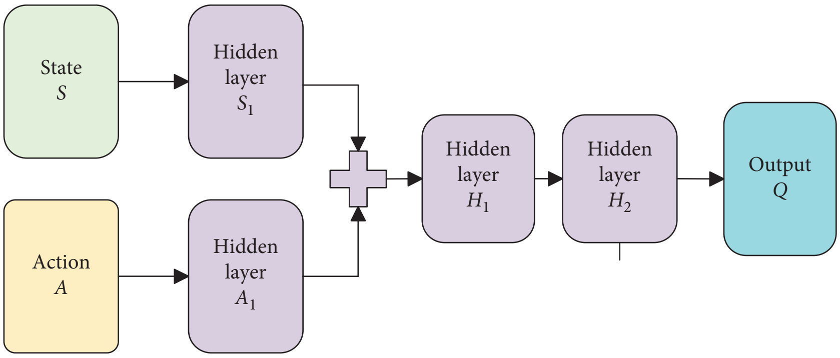

Figure 5: Structure of critic network and its target network.

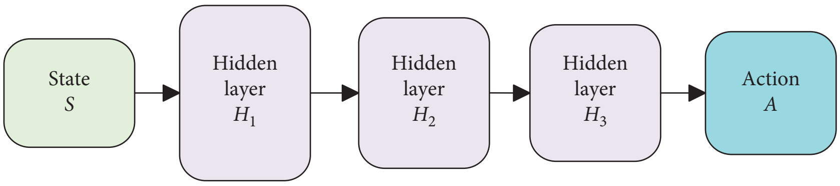

Figure 6: Structure of actor network and its target network.

Table 1: Final reward of each agent in different cases.

Option	Random	Perceptual learning ratio	Experience pools	Reward value	Training time (h)
1	×	1	1	0.162	21.6
2	√	1	1	0.194	19.54
3	√	30 : 1	1	0.226	13.65
4	√	30 : 1	2	0.23	12.16

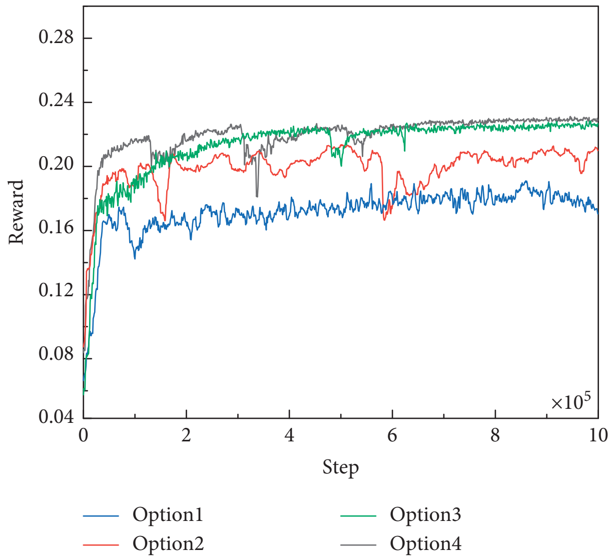

Figure 7: Reward curves in training process of different algorithms.

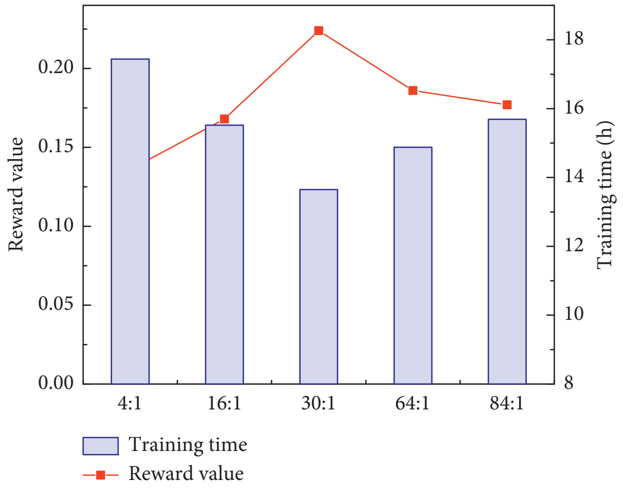

Figure 8: Relationship between training time and final reward value.

Table 2: Experimental results of each algorithm under different super parameters.

Greedy coefficient (E) ε	DQN	Actor-critic	PPO	DDPG
1.0–4	0.091	0.052	−0.152	0.191
5.0–5	0.123	0.075	−0.122	0.245
1.0–5	0.134	0.062	−0.091	0.217
5.0–6	0.139	0.109	−0.074	0.203

As can be seen from the table, under different greedy coefficients, the final reward values of each algorithm are also different, showing different trends. The PPO algorithm has the worst adaptability to the model, and the overall learning effect is the worst. The DQN algorithm is better than actor-critic algorithm, but it lags behind the proposed algorithm. In addition, with the decreasing of the greedy coefficient, the learning effect of each algorithm is improved. Actor-critic algorithm and DDPG algorithm need the appropriate greedy coefficient to achieve the optimal learning effect.

From Table 3, we can see that the algorithms exhibit different performance with the change of exploring coefficient in the strategy of random exploration. On the whole, each algorithm can use the better learning effect between 0.5 and 0.75; the higher the exploring coefficient, the lower the randomness of the searching algorithm, and it is easy to fall into the local convergence, leading to insufficient learning, poor learning effect, and low exploring coefficient, which will lead to too random and frequent exploration, not only seriously affecting the convergence time of the algorithm, but also leading to insufficient key information learning and poor learning effect.

In terms of training time and complexity of the algorithm, when setting the same hyperparameter, the comparison results of the four algorithms are shown in Table 4.

From the point of view of training complexity, the improved DDPG algorithm has a fast training speed due to the double-sample pool and reasonable environment perception-learning ratio and can keep good learning effect under the fast training speed.

3.6. Comparison of Test Results. Fix the trained neural network weights, and select the actual data of a typical day in winter for testing. See Figures 9–11 for the temperature, basic load, market electricity price information, and wind power output value of that day, and set the following comparative examples.

Case 1: Genetic algorithm is used to test the typical daily data; Case 2: The traditional DDPG algorithm is used to test the typical daily data (the algorithm corresponds to scheme 1 in Section 3.1); Case 3: The improved DDPG algorithm proposed in this paper is used to test the typical day (algorithm comparison, Section 3.1, scheme 4).

In the above examples, the result obtained by Case 1 can be considered as the optimal control result. The total economic benefits of intelligent parks and the calculation time of algorithms under different examples are shown in Table 5.

As can be seen from Table 5, under three examples, the total income of the park obtained by genetic algorithm is the highest, but at the same time, its calculation time is also the longest. The final benefit values of the two schemes using deep reinforcement learning algorithm are quite different. After training based on the traditional DDPG algorithm, the final test effect is poor, and the total income value is only 339.19, far lower than the genetic algorithm and the improved DDPG algorithm proposed in this paper. In addition, it is worth noting that deep reinforcement learning is essentially a model-free algorithm. It can learn decision-making experience from massive historical data and continuously improve the decision-making ability of agents, although the final test results are not as good as the traditional model-based optimization algorithm. However, the improved DDPG algorithm can greatly reduce the calculation time and even realize the decision-making ability of millisecond response, which cannot be realized by the

Table 3: Experimental results of each algorithm under different super parameters.

Exploring coefficient P	DQN	Actor-critic	PPO	DDPG
0.05	—	0.043	−0.922	0.153
0.25	—	−0.024	−0.089	0.197
0.5	—	0.075	−0.922	0.245
0.75	—	0.062	−0.922	0.213
0.95	—	0.014	−0.958	0.128

Table 4: Experimental results of each algorithm under different super parameters.

Attribute	DQN	Actor-critic	PPO	DDPG
Reward value	0.134	0.062	−0.091	0.217
Training time (h)	20.72	18.19	15.37	12.16

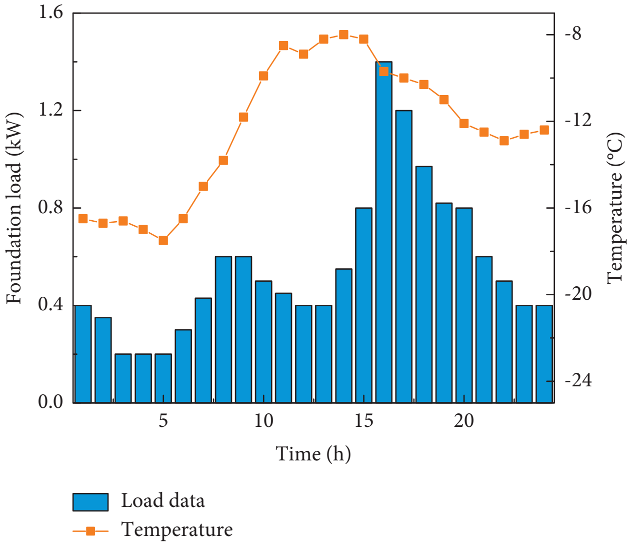

Figure 9: Load curve and temperature curve.

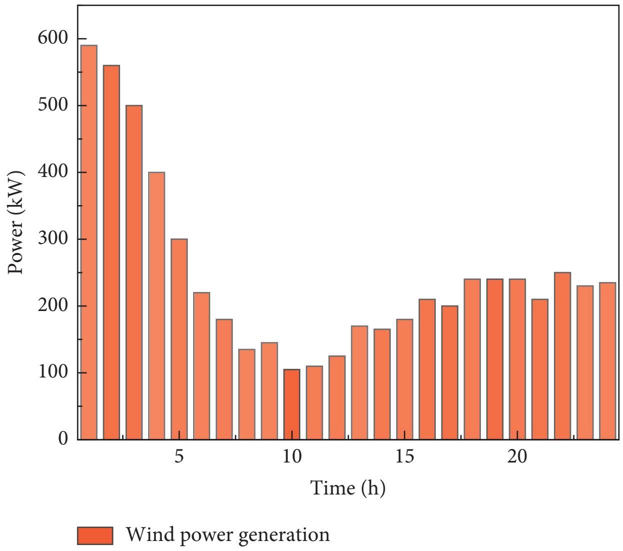

Figure 10: Wind power output result.

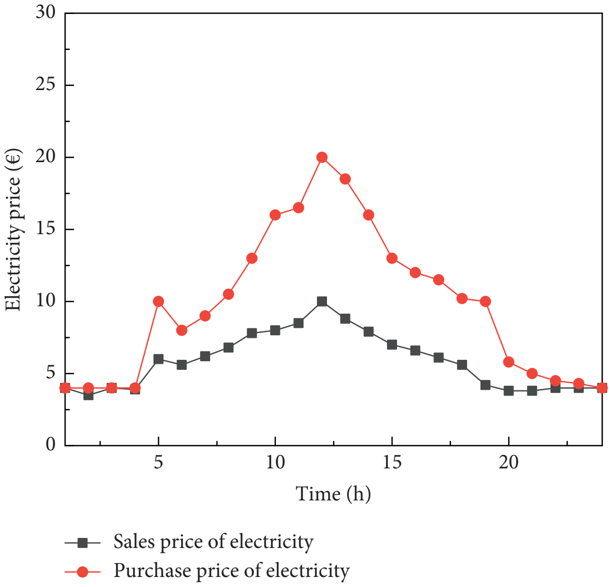

Figure 11: Electricity purchase price and electricity sale price curve.

Table 5: Total benefit of the park and the calculation time of the algorithm in different cases.

Parameter	Case 1	Case 2	Case 3
Profit (€)	428.35	339.19	413.26
Computing time (s)	5.89	0.065	0.039

traditional model-based solution algorithm, on the premise that the calculation results are not much different from the optimal control algorithm. In order to prove this conclusion, under the same test environment, the scale of solving the problem is further expanded. The comparison of computing time of three algorithms and the other DRL methods is shown as Table 6.

It can be seen that, with the continuous expansion of the scheduling cycle, the difference in solution time of the three algorithms is more obvious. Because of the offline learning and online decision-making ability of deep reinforcement learning, the calculation time basically increases linearly with the scheduling cycle, while the calculation time of traditional intelligent algorithms increases exponentially with the expansion of the scheduling cycle. With the increasing number of distributed new energy and controllable load connected to the intelligent park in the future, the overall dispatching scale of the park will gradually expand in the future, and the advantages of deep reinforcement learning algorithm will be further reflected at this time.

In addition, by comparing the test results of different algorithms, under different scheduling cycles, the target function values are shown respectively in Table 7.

As can be seen from the table, the test results of each algorithm are basically the same as the training results. The improved DDPG algorithm has the highest profit value in all the test algorithms because of its good learning effect in the training stage, and the PPO algorithm also has the lowest profit value in the test stage because of its poor learning effect.

Table 6: Calculation time of three algorithms under different scheduling period.

Scheduling period (h)	Case 1	Case 2	Case 3	DQN	PPO
24	5.89	0.065	0.039	0.041	0.038
168	216.56	0.536	0.156	0.176	0.164
720	1967.25	2.68	1.564	1.479	1.625

Table 7: Profit result of three algorithms under different scheduling period.

Scheduling period (h)	DQN	Actor-critic	PPO	DDPG
24	358.24	319.27	148.87	413.26
168	2301.87	2008.79	1374.43	2668.82
720	10820.17	9776.32	7876.91	12208.32

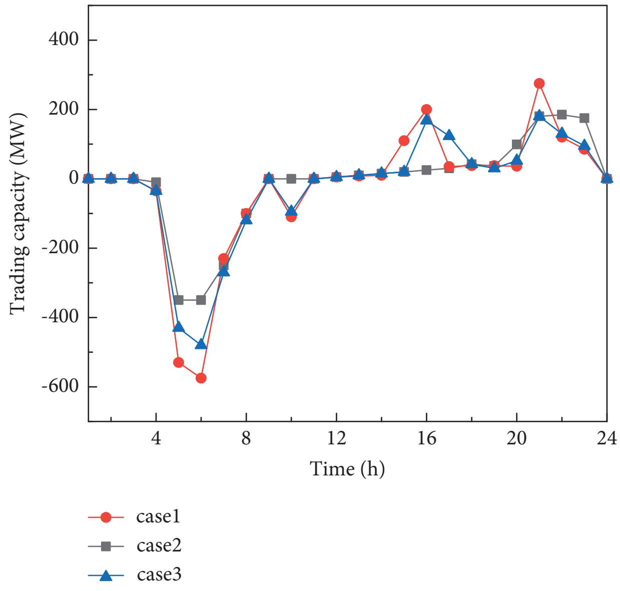

Figure 12: The results of market trading electricity under different algorithms.

Figure 12 shows the optimization results of interactive power consumption between intelligent park and superior power grid under different examples. It can be seen from the figure that, under the problem framework of this paper, EMS system chooses to sell electricity to the superior power grid in 4–10 h when the market electricity price is higher and purchases electricity from the superior power grid in 16–22 h when the market electricity price is lower, which not only meets its aggregate load resource electricity demand, but also maximizes its operating income. In addition, although both the traditional DDPG algorithm and the improved DDPG algorithm can keep the same change level with the optimal results, the improved DDPG algorithm is obviously closer to the optimal results in numerical results, and the control effect is better than the basic DDPG algorithm.

In addition, in order to reflect the impact of load side regulation on the overall income of the park, the following comparative examples are set up, respectively:

Table 8: Test result of three algorithms under different scheduling period.

Scheduling period	Case 1	Case 2	Case 3	Case 4
Profit (€)	413.26	379.27	356.99	298.76
Trading electricity (kW)	2601.89	2508.66	1887.38	1765.62

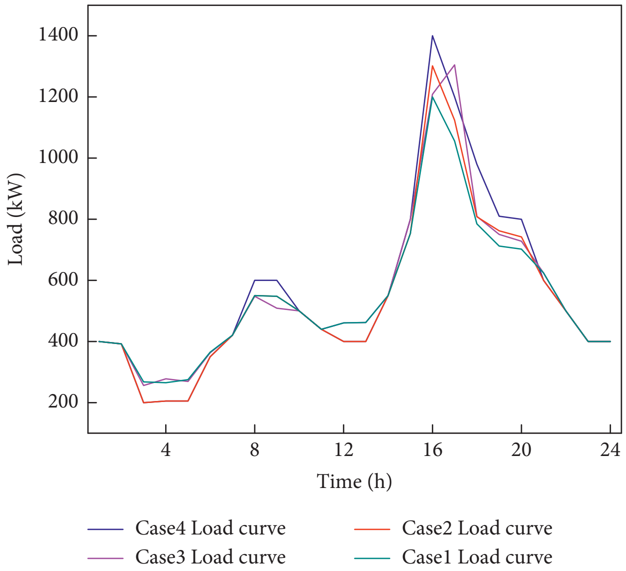

Figure 13: Equivalent load curve under different examples.

Case 1: consider regulating price-based demand response resources and temperature-controlled load resources simultaneously

Case 2: only consider regulating price-based demand response resources

Case 3: only consider regulating temperature-controlled load resources

Case 4: do not consider any load side resources

Under four scenarios, the results of revenue and total transactions with the grid under the 24-hour scheduling cycle are shown in Table 8.

From the results, it can be concluded that considering user-side resources in the form of demand response can effectively reduce the operation cost of the park. By responding positively in the period of higher electricity price and buying a large amount of electricity in the period of low electricity price to store energy for temperature-controlled load, the load curve of the park can be improved greatly while improving the economic benefit of the park. The specific equivalent load curve is shown in Figure 13.

4. Conclusion

In this paper, an energy management model of intelligent park considering the price-based demand response load and temperature-controlled load on the user side is established, and a model-free deep reinforcement learning algorithm is used to solve the problem. After analysis and verification, the following conclusions are drawn:

(1) Based on the traditional DDPG algorithm, the convergence speed of the algorithm and the final reward value can be improved by random exploration strategy, changing the perceptual learning ratio and increasing the high admission experience pool to a certain extent.

(2) Appropriate perceptual learning ratio can not only greatly reduce the training time of the model, but also avoid the algorithm falling into local optimal solution. However, the setting of perceptual learning ratio should be moderate, and too large or too small perceptual learning ratio is not conducive to the improvement of training effect.

(3) The model-free reinforcement learning algorithm can realize offline learning and online application. After the model is trained and converged, it can be applied to the online scheduling decision of intelligent park, and the calculation time is much less than that of the model-based solution algorithm.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

This work was sponsored in part by Natural Science Research Projects of Universities in Anhui Province, China (KJ2020A0968), and The Universities Outstanding Youth Projects of Anhui Province, China (Grant no. gxyq2020042).

References

[1] J. Su, R. Xu, S. Yu, B. Wang, and J. Wang, “Idle slots skipped mechanism based tag identification algorithm with enhanced collision detection,” *KSII Transactions on Internet and Information Systems*, vol. 14, no. 5, pp. 2294–2309, 2020.

[2] X. Ning, K. Gong, W. Li, L. Zhang, X. Bai, and S. Tian, “Feature refinement and filter network for person re-identification,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 9, pp. 3391–3402, 2021.

[3] G. Chen, Q. Pei, and M. M. Kamruzzaman, “Remote sensing image quality evaluation based on deep support value learning networks,” *Signal Processing: Image Communication*, vol. 83, Article ID 115783, 2020.

[4] W. Qin, Y. Hu, JunLei, and Y. Wang, “Comfort design and optimization of direct expansion multi-connected radiant air conditioning based on 3D flow field simulation,” *Displays*, vol. 69, Article ID 102054, 2021.

[5] Y. Wang, Le Sun, and S. Subramani, “CAB: Classifying arrhythmias based on imbalanced sensor data,” *KSII Transactions on Internet & Information Systems*, vol. 15, no. 7, pp. 2304–2320, 2021.

[6] L. Zhang, W. Li, L. Yu, L. Sun, X. Dong, and X. Ning, “GmFace: an explicit function for face image representation,” *Displays*, vol. 68, no. 1, Article ID 102022, 2021.

[7] A. J. D. Rathnayaka, V. M. Potdar, T. S. Dillon, O. K. Hussain, and E. Chang, “A methodology to find influential prosumers in prosumer community groups,” *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 706–713, 2014.

[8] A. C. Luna, N. L. Diaz, M. Graells, J. C. Vasquez, and J. M. Guerrero, “Cooperative energy management for a cluster of households prosumers,” *IEEE Transactions on Consumer Electronics*, vol. 62, no. 3, pp. 235–242, 2016.

[9] C. Wang and Li Peng, “Development and challenges of distributed generation, microgrid and intelligent distribution network,” *Power system automation*, vol. 34, no. 02, pp. 10–14, 2010.

[10] L. A. N. Zhu, Y. Zheng, X. Yang, and Y. Fu, “Integrated resource planning method for microgrid with demand side response,” *Chinese Journal of electrical engineering*, vol. 34, no. 16, pp. 2621–2628, 2014.

[11] H. Zhang, X. Shen, H. Mu, A. Liu, and h. Wang, “On line optimal scheduling of residential energy consumption based on multi agent asynchronous deep reinforcement learning,” *Chinese Journal of electrical engineering*, vol. 40, no. 1, pp. 117–127+379, 2020.

[12] F. Valencia, J. Collado, D. Marin, and L. G. Marín, “Robust energy management system for a microgrid based on a fuzzy prediction interval model,” *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1486–1494, 2016.

[13] Y. Zhang, N. Gatsis, and G. B. Giannakis, “Robust energy management for microgrids with high-penetration renewables,” *IEEE Transactions on Sustainable Energy*, vol. 4, no. 4, pp. 944–953, 2013.

[14] G. Sun, W. Qian, W. Huang et al., “Stochastic adaptive robust dispatch for virtual power plants using the binding scenario identification approach,” *Energies*, vol. 12, no. 10, p. 1918, 2019.

[15] D. E. Olivares, C. A. Canizares, and M. Kazerani, “A centralized energy management system for isolated microgrids,” *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1864–1875, 2014.

[16] B. Zhang, Q. Li, L. Wang, and W. Feng, “Robust optimization for energy transactions in multi-microgrids under uncertainty,” *Applied Energy*, vol. 217, pp. 346–360, 2018.

[17] D. Lei, S. Tu, Li Ye, and T. Pu, “Master slave game based dynamic pricing and energy management of multi virtual power plants,” *Power grid technology*, vol. 44, no. 3, pp. 973–983, 2020.

[18] Y. Zhao, “Ai Xin. Distributed optimal dispatch of integrated energy buildings considering power sharing,” *Power grid technology*, vol. 44, no. 10, pp. 3769–3778, 2020.

[19] R. Tao, G. Li, q. Wang, c. Hu, J. Zhang, and J. Ding, “Nash bargaining method for day ahead energy trading of multi microgrid on distribution side,” *Power grid technology*, vol. 43, no. 7, pp. 2576–2585, 2019.

[20] A. Askarzadeh, “A memory-based genetic algorithm for optimization of power generation in a microgrid,” *IEEE Transactions on Sustainable Energy*, vol. 9, no. 3, pp. 1081–1089, 2016.

[21] D. Wu, C. Gao, and Z. Ji, “Application of hybrid particle swarm optimization algorithm in economic optimal operation of microgrid,” *Control theory and application*, vol. 35, no. 4, pp. 457–467, 2018.

[22] C. She, C. Sun, Z. Gu et al., “A tutorial on ultrareliable and low-latency communications in 6G: integrating domain

knowledge into deep learning," *Proceedings of the IEEE*, vol. 109, no. 3, pp. 204–246, 2021.

[23] Y. Su, X. Lu, Y. Zhao, L. Huang, and X. Du, "Cooperative communications with relay selection based on deep reinforcement learning in wireless sensor networks," *IEEE Sensors Journal*, vol. 19, no. 20, pp. 9561–9569, 2019.

[24] J. Cao, D. Harrold, Z. Fan, T. Morstyn, D. Healey, and K. Li, "Deep reinforcement learning-based energy storage arbitrage with accurate lithium-ion battery degradation model," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4513–4521, 2020.

[25] H. Li, Z. Wan, and H. He, "Constrained EV charging scheduling based on safe deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2427–2439, 2020.

[26] D. Qiu, Y. Ye, D. Papadaskalopoulos, and G. Strbac, "A deep reinforcement learning method for pricing electric vehicles with discrete charging levels," *IEEE Transactions on Industry Applications*, vol. 56, no. 5, pp. 5901–5912, 2020.

[27] B. Wang, Y. Li, W. Ming, and S. Wang, "Deep reinforcement learning method for demand response management of interruptible load," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3146–3155, 2020.

[28] S. Bahrami, Y. C. Chen, and V. W. S. Wong, "Deep reinforcement learning for demand response in distribution networks," *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 1496–1506, 2021.

[29] Y. Ye, D. Qiu, J. Li, and G. Strbac, "Multi-period and multi-spatial equilibrium analysis in imperfect electricity markets: a novel multi-agent deep reinforcement learning approach," *IEEE Access*, vol. 7, pp. 130515–130529, 2019.

[30] Y. Liu, D. Zhang, and H. B. Gooi, "Optimization strategy based on deep reinforcement learning for home energy management," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 3, pp. 572–582, 2020.

[31] Y. Li, W. Zheng, and Z. Zheng, "Deep robust reinforcement learning for practical algorithmic trading," *IEEE Access*, vol. 7, pp. 108014–108022, 2019.

[32] T. Yang, L. Zhao, and W. Zomaya, "Dynamic energy dispatch strategy for integrated energy system based on improved deep reinforcement learning," *Energy*, vol. 235, no. 15, Article ID 121377, 2021.

[33] L. Peng, y. Zhang, J. Xu, S. Liao, and L. Yang, "Adaptive uncertain economic dispatch based on deep reinforcement learning," *Power system automation*, vol. 44, no. 9, pp. 33–46, 2020.

[34] G. Chen, L. Wang, M. Alam et al., "Intelligent group prediction algorithm of GPS trajectory based on vehicle communication," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 3987–3996, 2020.

[35] I. Kamwa, S. R. Samantaray, and G. Joos, "On the accuracy versus transparency trade-off of data-mining models for fast-response PMU-based catastrophe predictors," *IEEE Transactions on Smart Grid*, vol. 3, no. 1, pp. 152–161, 2012.

[36] T. A. Nakabi and P. Toivanen, "An ANN-based model for learning individual customer behavior in response to electricity prices," *Sustainable Energy, Grids and Networks*, vol. 18, no. C, Article ID 100212, 2019.