International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
Impact Factor (2024): 7.741

Electric Energy Demand Response Algorithm based on Deep Reinforcement Learning

Yun Ju, Weixing Gao, Xiaoqi Shao

North China Electric Power University, School of Control and Computer Engineering,
No. 2 Beinong Road, Changping District, Beijing, China
474754536@qq.com

Abstract: ***With the rapid development of smart grid and renewable energy, demand response mechanism has become a key means to balance microgrid supply and demand and improve energy efficiency. As the traditional power demand response algorithm is difficult to deal with the uncertainty of power demand and the adverse effects caused by customers, this paper proposes a power demand response algorithm based on deep reinforcement learning. Firstly, we study and design a price - based layered electricity energy demand response model. Then, the dynamic optimal pricing decision of power trading market is described as Markov decision process, and the learning mechanism under the framework of deep reinforcement learning is expounded. Finally, a solution algorithm based on deep reinforcement learning is designed. The simulation results show that the proposed algorithm can adjust the electricity price adaptively according to the load demand, which can reduce the cost of users by about 17% and reduce the peak load demand.***

Keywords: Demand response, Deep reinforcement, Smart grid, electric energy

1. Introduction

With the development of social productivity, electric energy has become an indispensable existence for people. The traditional power system mainly relies on thermal power generation. By dispatching conventional generator sets, the power load curve is tracked, the instantaneous balance of system power is maintained, and the source is actuated with load. However, with the rapid expansion of electricity demand, the continuous shortage of fossil fuel resources, and the combustion of fossil fuels in thermal power generation produces a large number of harmful gases, resulting in environmental pollution and other negative effects. In recent years, under the policy of actively and steadily promoting carbon peak and carbon neutrality in China, with the access of new energy such as photovoltaic and wind power with a high proportion, due to the intermittent nature of new energy, it is difficult to adapt to the traditional way of dispatching conventional generator sets, and there is an urgent need for a new type of power system to realize the collaborative interaction of transmission, distribution, utilization and storage, and realize the revolutionary transformation from traditional power grid to smart grid [1]. In recent years, the concept of demand-side management (DSM) has attracted a lot of attention in smart grids [2]. As a typical approach to DSM, demand response (DR) is widely seen as the most cost-effective and reliable solution for improving the efficiency and reliability of power systems [3].

DR Refers to when the reliability of the power system is threatened or the price of electricity rises, the user receives the incentive information of the power load usage process or the signal of the retail price rise. By changing their consumption habits and reducing or delaying the power load during peak hours, the user can improve energy efficiency, reduce user costs, reduce carbon emissions and improve the stability of the power grid [4].

2. Related work

At present, demand response algorithms are mainly studied from two categories: incentive-based and price-based. Customers who participate in incentive-based DR Programs can receive discounted retail prices or separate incentive payments for pre-signing or measured load reduction. Price-based DR, including time-of-use (TOU) rates [5] and real-time pricing (RTP) [6], refers to plans where customers respond to time-varying changes in retail electricity prices. Although both of these demand response mechanisms can promote active participation of loads, as described in [7], price-based demand response management is more common than incentive-based demand response research management, so this study focuses on price-based demand response.

The price-based DRM project has carried out a number of research efforts around the world [8], [9], and the energy consumption of household appliances is a major factor in price-based DRM programs. For example, systems based on mixed integer linear programming (MILP) have been designed to determine optimal equipment scheduling, thereby improving energy efficiency and reducing consumer costs [10,11].

Aiming at maximizing the profit of microgrid retailers, the author transforms the electricity price and microgrid scheduling problems into a mixed integer quadratic programming problem, and studies the dynamic pricing strategies of microgrid retailers in integrated energy systems [12].

Similarly, Yu and Hong [13] proposed a DRM method based on real-time price. Through Stackelberg game, the power retailer of the facility energy management center is established, and users purchase resources from it to achieve the optimal load control of the equipment, thus forming the optimal strategy.

Hande Yaman et al. [14] proposed a multi-stage stochastic programming model and established their own optimization equations for different models of typical scenes respectively. However, the optimization accuracy was limited when encountering more complex models and unexpected situations.

At present, most methods rely on traditional deterministic rules or abstract models that do not ensure optimality when dealing with unstable energy systems, have limited optimization accuracy when models are more complex and encounter difficult to predict situations, and game theory faces scalability problems due to large numbers of binary values when the system is large.

In order to solve the above problems, reinforcement learning (RL) is a prominent solution. By interacting with the random environment, the agent selects actions to make decisions to the environment, and the environment generates new states and rewards to the agent, and adaptively learns the best behavior, thus maximizing the cumulative rewards [15].

Literature [16] proposed the demand response of residential and small commercial buildings based on Q-learning, designed the uncertain load demand and power grid information as the state quantity of the system, and explored the optimization scheme.

Similarly, Q-learning algorithm is used in literature [17] to solve the price-based demand response problem of microgrid. However, when high-dimensional space is involved, Q-table method consumes a lot of storage space and computing resources, and its practical application has limitations.

Deep reinforcement learning (DRL) combines the decision-making ability of RL with the information perception ability of deep learning (DL), uses the generalization ability of DL itself to deal with the uncertainty of load demand, and solves the problem that RL involves continuous states and actions, which requires huge storage and computing resources, and obtains the optimal solution. A model-free algorithm to solve complex control problems [18].

In recent years, there have been some attempts to apply DRL algorithm to DR. a DRL-based energy management algorithm was developed in [19] to minimize the cost of electricity for smart homes, The uncertainties of the model and parameters are also considered.

The paper [20] designs a role-critic based DRL algorithm to determine the optimal energy management strategy for industrial facilities. All of these papers express DR Control as a Markov decision process (MDP) and use their respective DRL algorithms to make complex DR Decisions adapted to specific constraints.

Therefore, this paper studies the power demand response method based on DRL. The main contributions are as follows: 1) Considering the user's power load usage, a dynamic pricing DR Method for power demand response in hierarchical power market is proposed; 2) Deep reinforcement learning algorithm is proposed to illustrate the hierarchical decision-making framework, and the dynamic pricing of retail electricity price is expressed as a Markov decision process (MDP), and the optimal pricing is solved based on DDQN algorithm; 3) Solve the uncertainty of power grid load demand curve through online learning, and coordinate the impact of users' private preferences on the market through the unsatisfactory cost function.

The rest of this article is organized as follows. The first section describes the overall framework and MDP modeling. The second section introduces the demand response algorithm based on DRL in detail. The third section provides the numerical experimental results. Finally, the fourth section summarizes the thesis.

3. Electricity energy demand response system

This paper aims to construct a price-driven electricity energy demand response (EEDR) model based on electricity integrated energy market. This model is a microgrid system including grid operator, electricity service provider and multiple electricity users. As shown in Figure 1, in this demand environment, electricity service provider buys electricity from grid operator at wholesale electricity price, and then sells electricity to power users at retail electricity price. Users change their electricity demand according to their own load demand and the electricity price signal given by electricity service provider.

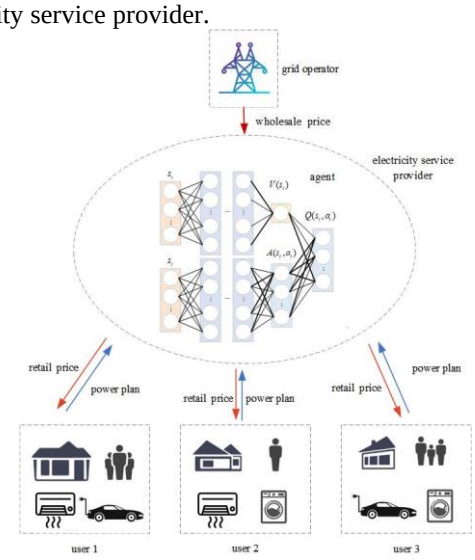

Figure 1: Hierarchical model of electricity

3.1 electricity service provider model

As a participant in the electricity wholesale market, the electricity service providers cooperate closely with the power grid to jointly maintain the stable operation of the market. At each specific time period, the electricity service provider purchases electricity at the wholesale price set by the grid, sets its own retail price on this basis, and sells the energy to the end electricity user. The core of this model is that electricity service providers can flexibly adjust retail prices according to market supply and demand conditions, electricity costs and other factors to meet the needs of different users.

The object function of the electricity service provider is to implement a dynamic determination of the retail price to maximize its profit, as shown below

$$\max \sum_{n=1}^{N}\sum_{t=1}^{T}\left(\lambda_{t,n}-\pi_t\right)\cdot\left(e_{t,n}^{\text{curt}}+e_{t,n}^{\text{critic}}\right) \quad (1)$$

Where, $\lambda_{t,n}$ represents the retail price of electricity, π_t represents the wholesale price of electricity, $e_{t,n}^{\text{curt}}$ represents the actual electricity consumption that can be reduced, and $e_{t,n}^{\text{critic}}$ represents the actual electricity consumption that cannot be reduced.

3.2 electricity user model

electricity users participating in demand response, after receiving the price signal issued by the electricity service provider, will make the corresponding power energy demand response decision based on the comprehensive consideration of economic cost and user satisfaction.

The objective function for user n is to minimize its overall cost, as described below

$$\min \sum_{t=1}^{T}\left[\lambda_{t,n}\cdot\left(e_{t,n}^{\text{curt}}+e_{t,n}^{\text{critic}}\right)+\varphi_{t,n}\right] \quad (2)$$

$\varphi_{t,n}$ Represents the unsatisfied cost of ginseng when the user reduces the energy demand for time period t, the calculation is as follows:

$$\varphi_{t,n}=\frac{\alpha_n}{2}\left(E_{t,n}^{curt}-e_{t,n}^{curt}\right)^2+\beta_n\left(E_{t,n}^{curt}-e_{t,n}^{curt}\right) \quad (3)$$

Where α_n and β_n is the user dependent parameter, α_n representing the preference parameter of user n for reducing energy load, and its range is between [0,1]. This means that if the value of α_n is larger, the user will be unwilling to make a higher price reduction, and the user's satisfaction will be higher. β_n is the default parameter for the cost of user dissatisfaction, with a value between {0,1}. $E_{t,n}^{curt}$ indicates that the user can reduce the expected consumption of load.

The utility function for reducing energy load is defined as follows [4]:

$$\mathrm{e}_{t,n}^{curt}=E_{t,n}^{curt}(1+\xi_t\cdot\frac{\lambda_{t,n}-\pi_t}{\pi_t}) \quad (4)$$

Where, the elastic coefficient of time period t is ξ_t

3.3 model objective function

In this paper, the objective function is the balance between the electricity profit of the electricity service provider and the electricity cost of the user. The formula is as follows:

$$\max \sum_{n=1}^{N}\sum_{t=1}^{T}\left[\rho\cdot\left(\lambda_{t,n}-\pi_t\right)\cdot e_{t,n}+(1-\rho)\cdot\left(\lambda_{t,n}\cdot e_{t,n}+\varphi_{t,n}\right)\right] \quad (5)$$

$$e_{t,n}=e_{t,n}^{curt}+e_{t,n}^{\text{critic}} \quad (6)$$

Where, ρ indicates that the weight factor ranges between (0,1) to represent the relative importance of the user's electricity energy bill and the profit of the electricity service provider.

4. DRL-based optimal dynamic tariff model

4.1 model objective function

The basic idea of RL is to learn the optimal strategy to maximize the cumulative reward value or achieve a specific goal through the interaction between agents and the environment [21]. In RL solving problems, the environment is usually normalized as a Markov decision process (MDP). MDP is a sequential mathematical model composed of three basic elements: state, action and reward. Its characteristics can be understood as that the action taken by the agent in the current state not only affects the current feedback, but also affects the next state and feedback [21].

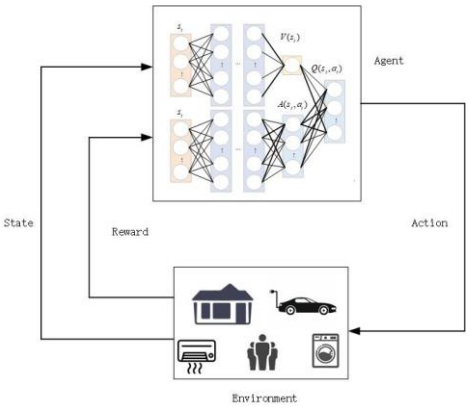

Figure 2: The overall diagram of the proposed MDP

In this paper, the dynamic retail pricing problem is first expressed as MDP, as shown in Figure 2, in which the electricity service provider acts as an agent, the electricity user represents the environment, the retail price represents the action the service provider sends to the user at every moment, the energy information of the user (energy demand and consumption) represents the state, and the profit of the electricity service provider and the cost of the user represent the reward.

The action space of the agent is mainly the retail electricity price $\lambda_{t,n}$ set by the electricity selling service provider to the user n at different times. The upper and lower bounds of the retail electricity price are $\lambda^{\max}$ and $\lambda^{\min}$ respectively.

$$\lambda^{\min}<\lambda_{t,n}<\lambda^{\max}$$

The agent reward function is used to help the agent judge whether the selection action is good or bad. The goal is to maximize the profit of the electricity service provider and the user cost. The measurement index is the maximum cumulative reward that can be obtained from the environment designed around this goal

$$r\left(e_{t,n}|\ E_{t,n},\lambda_{t,n}\right)=\sum_{n=1}^{N}\left[\rho\cdot\left(\lambda_{t,n}-\pi_t\right)\cdot e_{t,n}-(1-\rho)\cdot\left(\lambda_{t,n}\cdot e_{t,n}+\varphi_{t,n}\right)\right] \quad (7)$$

4.2 solving algorithm based on Dueling DQN

After defining the state, action and reward functions in the above section, the goal of the power energy demand response system is to find the control action to maximize the profit of the power supplier and the cost of the user, and the dynamic optimal price. This paper determines the optimal pricing method for the power energy system based on the DDQN algorithm.

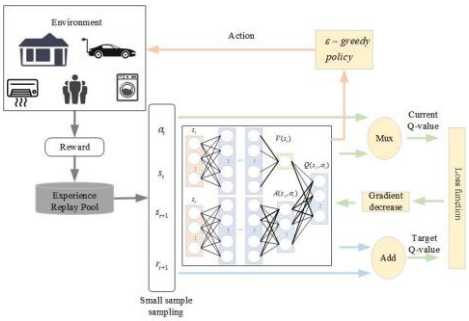

Figure 3: Algorithmic model based on DDQN

The value function approximation process based on DDQN is shown in Figure 3. The input of DDQN and DQN is the same, both of which are state information, but the output is different. Traditional DQN only includes a state action estimation network, while the output of Dueling algorithm includes state value V and the advantage value of each action. Finally, the action value of each action can be obtained by combining the state value with the advantage value through the depth model. The significant difference between DDQN and DQN is the structure of the estimated neural network. Unlike traditional DQN, which contains only one state action estimation network, DDQN architecture represents both the state value network $V(s_t)$ and the action dominance network $A(s_t, a_t)$, and uses a single depth model whose output combines the two to produce the state-action value $Q(s_t, a_t)$. The Q function based on DDQN structure is defined as (10):

$$Q(s_t, a_t) = V(s_t) + \left[A(s_t, a_t) - \frac{1}{|\mathrm{A}|} \sum_{a_t' \in \mathrm{A}} A(s_t, a_t') \right] \quad (8)$$

Where A is the set that contains all the executable actions, and |A| is the number of all the executable actions. The action advantage function is set as a single action function minus the average value of all action advantage functions in a certain state to eliminate redundant degrees of freedom and improve the stability of the algorithm.

The pseudo-code of the solution algorithm based on DDQN is shown in the table. The algorithm can be decomposed into three stages: initialization (lines 1~2), experience accumulation (lines 4~10), and experience learning (lines 12 ~ 18). During initialization, set the hyperparameters of the DRL algorithm. Then, the DDQN is initialized with the random parameter w, and the experience pool is initialized to the empty set. Starting with line 3, the algorithm enters scenario iteration. At the beginning of each episode, the initial state is randomly reset to remove the coupling between the sample and time during the learning process. The algorithm is experienced from lines 4 to 10. The step counter t is increased in detail, the actions are selected according to the ε-greedy strategy, and the state action transition tuples are successively stored in the experience pool. When the number of samples in the pool accumulates beyond the replay start size M, the experiential learning process is performed from line 12 to line 18. Specifically, take a random batch of samples numbered n from the pool in line 12. Then, the target Q value and the predicted Q value of the sample are calculated respectively in line 13, on which the loss function is calculated as shown in line 15. Finally, in line 16, the weights in the DDQN are updated using the batch gradient Descent (BGD) method.

Algorithm 1: Demand response algorithm of electric energy management based on DDQN

1: Randomly initialize DDQN parameters ω
2: Initialize replay buffer
3: **for** each epoch **do**
4: observe s_t ;
5: **for** each time **do**
6: choose an action a_t using the $\varepsilon - greedy$
7: obtain reward a_t , and observe the next state s_{t+1}
8: **If** sample size > N
9: remove the oldest observation sample
10: store (s_t, a_t, r_t, s_{t+1}) into the experience pool;
11: **If** sample size > M
12: sample random mini-batch of (s_t, a_t, r_t, s_{t+1}) with number n from experience pool
13: obtain the target Q values and the predicted Q values respectively
14: calculate the loss function
15: update the parameters ω by BGD method
16: **end for**
17: **end for**

5. Analysis of Experimental Results

The previous section introduced the overall technical method, and this section will give the numerical simulation results to evaluate the performance of the dynamic pricing demand response algorithm. For ease of explanation, this section conducts simulation on the basis of one electricity supplier and three power users to verify the effectiveness of the algorithm. In this paper, one power user is selected to describe the performance of the experiment.

5.1 example and input data

Sample load demand curves of the power grid at various periods were obtained from SDG&E [] as input in the flowchart shown in the experiment. The entire event cycle was divided into 24 time intervals, representing 24 hours of a day. Figure 3 shows the load demand graph of a power user: it shows the critical load and the adjustable load of the user. The critical load is the load that the user must use daily, which does not change with the change of the electricity price, and the adjustable load changes with the change of the electricity price.

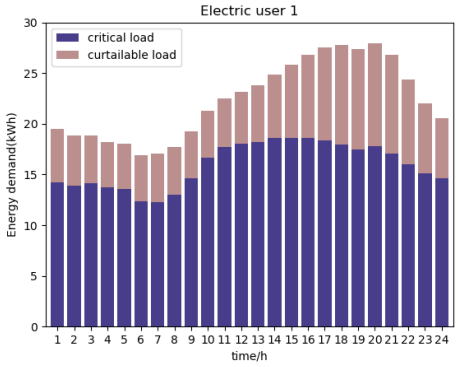

Figure 4: User load information

The user comfort correlation coefficient α_n and β_n of user 1 is set to 0.8 and 0.1 respectively. The elasticity index is shown in Table 1, with different responses during off-peak/mid-peak/peak hours.

Table 1: Load elastic coefficient

	off-peak	*mid-peak*	*ghts*
ξ_t	-0.3	-0.5	-0.7

Figure 5 shows the online data of wholesale electricity price shown by ComEd. The range of retail electricity price is represented by a certain wholesale electricity price coefficient. The weight coefficient ρ in this study is assumed to be 0.9, indicating that the profit of the electricity selling service provider has greater relative importance than the cost of the user.

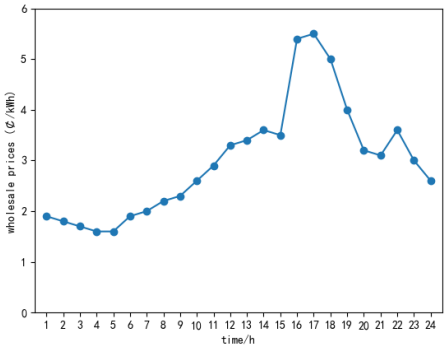

Figure 5: Wholesale electricity prices

The summary of the hyperparameters of algorithm 1 used in the simulation is shown in Table 2, and other hyperparameters related to neural networks are determined according to the conventions recommended by the deep learning community.

Table 2: Algorithm hyperparameter

hyperparameter	*value*
discount factor	0.7
learning rate	0.01
small batch size	30
learning steps	3000
experience pool	1000
restart size	100

5.2 electricity price optimisation results and analysis

After running the simulation, the main output is the optimal retail price for the power user. Figure 6 shows the optimal retail price and wholesale price signals for the three cu's, as well as the energy demand and actual energy consumption for the adjustable load. Because critical load demand does not change with retail prices, only adjustable load energy information is displayed.

As can be seen from Figure 6, the trend in retail prices is similar to the trend in wholesale prices, reflecting the cost of purchasing electric energy from the grid; From time period 6 to time period 12, the retail price per user increases in order to make more profit for the service provider selling electricity, but a sudden decline is observed at time period 14. This is because at time period 14, the elasticity coefficient changes from −0.3 to −0.5 to reflect the mid-peak period, and a sustained increase in retail prices will lead to a greater reduction in energy during this period. Compared with off-peak hours, the electricity price gap in peak hours is smaller than that in off-peak hours, but the energy consumption reduction gap (energy demand-energy consumption) is larger. This is because electricity demand is more elastic during peak hours and greater energy reductions can be achieved.

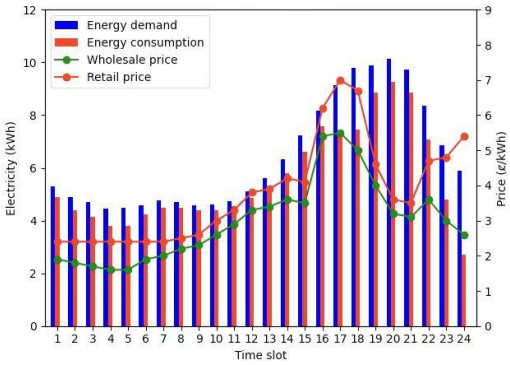

Figure 6: Algorithm optimisation results

5.3 verification of convergence of the algorithm

When PyTorch, an open-source Deep learning framework developed by Facebook, was used to evaluate the performance of neural networks, the loss functions of DDQN (Dueling Deep Q-Network) and traditional DQN (Deep Q-Network) structures were compared during iterative training. The comparison results are shown in Figure 7. Compared with DQN, the loss function using DDQN structure decreases faster and eventually reaches a smaller value. At the same time, the fluctuation of the loss function is smaller, indicating that the algorithm using DDQN is more stable.

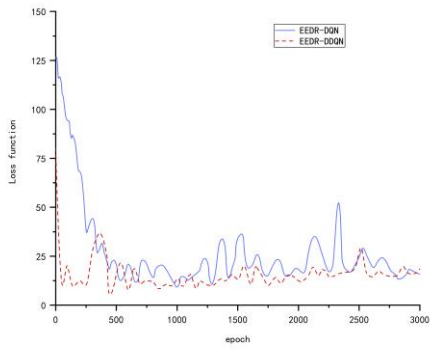

Figure 7: Loss function values for different algorithms

In order to prove the convergence of the proposed algorithm, the cumulative reward comparison between DDQN and DQN in the iterative training process is shown in Figure 8. It can be clearly seen from Figure 8 that the agent does not know how to select an action to obtain a high q value at the beginning, but with the progress of iteration, the cumulative reward gradually increases with the continuous trial-and-error learning of the power selling service provider from the environment, and finally converges to the maximum value. Compared with DQN, the maximum cumulative reward of DDQN is much larger around the 500th time, and the optimal strategy is generated, that is, the optimal behavior with the maximum cumulative reward is selected.

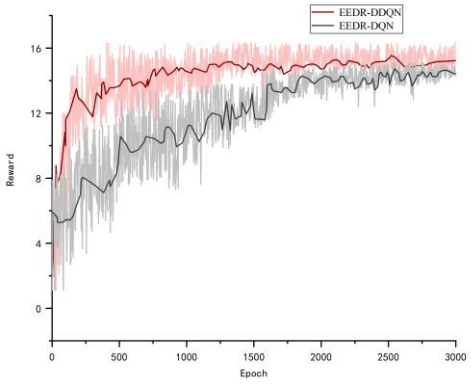

Figure 8: Cumulative rewards for different algorithms

Although there are some fluctuations in the training process due to the random action of the ε-greedy strategy, the overall trend of the curve proves the convergence of the algorithm

5.4 validation of the effectiveness of the algorithm

Figure 9 shows the total energy consumption reduction of each user after adding the dynamic pricing DR Algorithm proposed in this paper, and the green part represents the difference of the cumulative payment of the user's actual energy consumption before and after the demand response. As shown in Figure 9, user 1 reduced energy consumption by approximately 18%. Therefore, demand response provides an opportunity for power market to balance energy supply and demand, which can effectively eliminate system overload and improve the reliability of power system

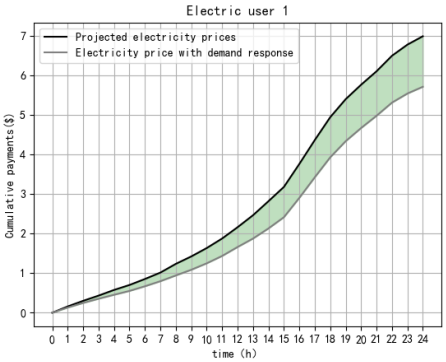

Figure 9: Cumulative payments by user

In order to study the effect of the weight factor ρ, we simulate the change of ρ between 0 and 1. Figure 10 and Figure 11 show the average retail electricity price and average profit of the RHO coupling of the service provider and the user, respectively. From these two graphs, we can observe that an increase in ρ from 0 to 1 leads to an increase in the average retail price and the average profit of the service provider selling electricity; However, the average profit of the league is down.

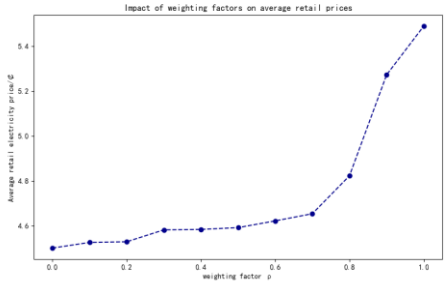

Figure 10: Impact of weighting factors on retail price

Because as ρ increases, the profit of the selling service provider becomes more important relative to the cost of the user. Especially in the case of ρ = 1, the electricity sales service provider aims to maximize its own profit, regardless of the cost of the user, so the electricity sales service provider chooses a relatively high retail price. In contrast, when ρ=0, the system tends to minimize the cost to the user; Therefore, the electricity service provider chooses a relatively low retail price to the user.

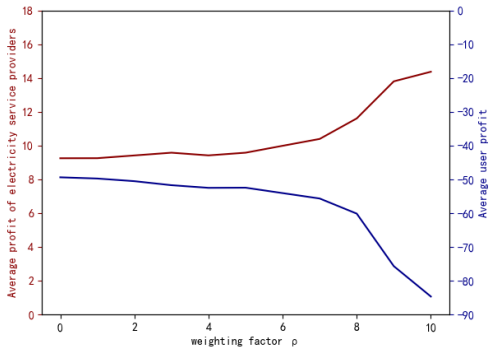

Figure 11: Impact of weighting factors on profits

6. Conclusion

With the rapid development of power grid construction and intelligent technology, the structure of electric power energy system is increasingly complex and diversified, and it is

necessary to balance supply and demand through electric power energy demand response to improve system stability. Traditional demand response strategies often operate based on preset rules or fixed feedback models, which to some extent limits their ability to adapt to complex and dynamic grid environments. Therefore, new, more flexible and adaptable energy demand response strategies need to be developed to ensure the stable operation of the power system and sustainable energy supply. Therefore, this paper proposes an electric energy demand response algorithm based on deep reinforcement learning DDQN. First, this chapter transforms the demand response problem into a Markov decision problem, then gives a complete definition of the demand response model with Markov properties, and then designs the EEDR-DDQN algorithm in detail. The contribution of this algorithm is as follows:

Therefore, a new power demand response algorithm based on deep reinforcement learning is proposed in this study. The main contributions are as follows:

(1) Considering the power load usage of users, a dynamic pricing DR Method of power energy demand response stratified power market is proposed.

(2) Deep reinforcement learning algorithm is used to illustrate the hierarchical decision-making framework, and the dynamic pricing of retail electricity price is expressed as a Markov decision process (MDP), and the optimal pricing is solved based on DDQN algorithm.

(3) Solve the uncertainty of power grid load demand curve through online learning, and coordinate the impact of users' private preferences on the market through the unsatisfactory cost function.

References

[1] L. Niamir, T. Filatova, A. V oinov, and H. Bressers, ‘‘Transition to low-carbon economy: Assessing cumulative impacts of individual behav-ioral changes,’’ Energy Policy, vol. 118, pp. 325–345, Jul. 2018, doi:10.1016/j.enpol.2018.03.045.

[2] C. Li, X. Y u, W. Y u, G. Chen, and J. Wang, “Efficient computation forsparse load shifting in demand side management,” IEEE Trans. Smart Grid, vol. 8, no. 1, pp. 250–261, Jan. 2017.

[3] J. S. V ardakas, N. Zorba, and C. V. V erikoukis, “A survey on demand response programs in smart grids: Pricing methods and optimization algorithms,” IEEE Commun. Surveys Tuts., vol. 17, no. 1, pp. 152–178,1st Quart., 2015.

[4] L. Park, Y. Jang, S. Cho, and J. Kim, “Residential demand response forrenewable energy resources in smart grid systems,” IEEE Trans. Ind.Informat., vol. 13, no. 6, pp. 3165–3173, Dec. 2017.

[5] Y. Hung and G. Michailidis, “Modeling and optimization of time-of-use electricity pricing systems,” IEEE Trans. Smart Grid, vol. 10, no. 4,pp. 4116–4127, Jul. 2019

[6] J. Qin, Y. Wan, X. Yu, F. Li, and C. Li, “Consensus-based distributed coordination between economic dispatch and demand response,” IEEE Trans. Smart Grid, vol. 10, no. 4, pp. 3709–3719, Jul. 2019.

[7] J. Qin, Y. Wan, X. Yu, F. Li, and C. Li, “Consensus-based distributed coordination between economic dispatch and demand response,” IEEE Trans. Smart Grid, vol. 10, no. 4, pp. 3709–3719, Jul. 2019

[8] S. Pal and R. Kumar, “Electric vehicle scheduling strategy in residential demand response programs with neighbor connection,” IEEE Trans. Ind.Informat., vol. 14, no. 3, pp. 980–988, Mar. 2018.

[9] R. Lu, S. H. Hong, and M. Y u, “Demand response for home energy management using reinforcement learning and artificial neural network,” IEEE Trans. Smart Grid, vol. 10, no. 6, pp. 6629–6639, Nov. 2019.

[10] N. G. Paterakis, O. Erdinc, A. G. Bakirtzis, and J. P. Catalão, “Optimal household appliances scheduling under day-ahead pricing and load-shaping demand response strategies,” IEEE Trans. Ind. Informat.,vol. 11, no. 6, pp. 1509–1519, Dec. 2015.

[11] A. Jindal, B. S. Bhambhu, M. Singh, N. Kumar, and K. Naik, “A heuristic-based appliance scheduling scheme for smart homes,” IEEE Trans. Ind. Informat., vol. 16, no. 5, pp. 3242–3255, May 2020.

[12] Jin M, Feng W, Marnay C, et al. Microgrid to enable optimal distributed energy retail and end-user demand response[J]. Applied Energy, 2018, 210: 1321-1335.

[13] M. Y u and S. H. Hong, “A real-time demand-response algorithm for smart grids: A Stackelberg game approach,” IEEE Trans. Smart Grid,vol. 7, no. 2, pp. 879–888, Mar. 2016.

[14] Şahin M K，Çavuş Ö，Yaman H．Multi-stage stochastic programming for demand response optimization[J]. Computers & Operations Research，2020，118：104928．

[15] V. Mnih et al., “Human-level control through deep reinforcement learning,” Nature, vol. 518, no. 7540, pp. 529–533, 2015

[16] WEN Zheng，O'Neill D，MAEI H．Optimal demand response using device-based reinforcement learning[J]．IEEE Transactions on Smart Grid，2015，6(5)：2312-2324．

[17] KIM B，ZHANG Y，SCHAAR M，et al. Dynamic pricing and energy consumption scheduling with reinforcement learning［J］

[18] V. Mnih et al., “Human-level control through deep reinforcement learn- ing,” Nature, vol. 518, no. 7540, pp. 529–533, 2015.

[19] L. Y u et al., “Deep reinforcement learning for smart home energy management,” IEEE Internet Things J., vol. 7, no. 4, pp. 2751–2762, Apr. 2020.

[20] X. Huang, S. H. Hong, M. Y u, Y. Ding, and J. Jiang, “Demand response management for industrial facilities: A deep reinforcement learning approach,” IEEE Access, vol. 7, pp. 82194–82205, 2019

[21] SUTTON R S，BARTO A G. Reinforcement learning：an introduction［M］.Cambridge，USA：MIT Press，2018.

Author Profile

Ju Yun received his bachelor's degree in Electronic Information Engineering from North China Electric Power University in June 2003, his master's degree in computer Science and technology from North China Electric Power University in April 2006, and his doctor's degree in Control theory and Control Engineering from North China Electric Power University in June 2014.His research interests include cloud computing, big data, Internet of Things, mobile Internet, artificial intelligence, information security, etc. Recently, he has focused on the application of power Internet of Things and artificial intelligence in energy Internet.

Weixing Gao received his bachelor's degree in computer Science and technology from Hebei University of Architecture in June 2021; he is studying for a master's degree in computer Science and technology from North China Electric Power University, His research interests include demand response and artificial intelligence in smart microgrids.

Xiaoqi Shao received his bachelor's degree in computer Science and technology from Jinzhong University in June 2021; he is studying for a master's degree in computer Science and technology from North China Electric Power University, His research interests include energy management optimization in microgrids and deep reinforcement learning.