Dynamic Pricing Strategy in Electricity Trading Market Based on Reinforcement Learning

line 1: Chunlin Hu
line 2: *Faculty of Electronic and Information Engineering*
line 3: *Xi'an Jiaotong University*
line 4: Xian 710049, China
line 5: hucl0918@stu.xjtu.edu.cn

line 1: Dou An
line 2: *Faculty of Electronic and Information Engineering*
line 3: *Xi'an Jiaotong University*
line 4: Xian 710049, China
line 5: douan2017@xjtu.edu.cn

2023 35th Chinese Control and Decision Conference (CCDC) | 979-8-3503-3472-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/CCDC58219.2023.10327004

***Abstract*—Intermediaries play an important role in electricity trading in the smart grid. However, smart grids are highly dynamic and complex due to diverse consumer demand, uncertain electricity market prices, and competition among different intermediaries. It brings a huge challenge for intermediaries to set reasonable electricity prices to maintain supply and demand balance and maximize their profits. This paper presents an electric energy trading model based on an intermediary. In this model, the electricity purchased by the user is negatively correlated with the retail price published by the intermediary. Intermediaries can make maximum profit by adjusting retail prices. Considering that deep reinforcement learning (DRL) can solve uncertain decision-making problems, this paper uses the deep Q-network (DQN) algorithm to develop real-time pricing strategies for intermediaries in the electricity market. Then, the validity of the pricing strategy of the intermediary is verified by the simulation of electricity consumption data in the real world. After comparison with non-learned fixed pricing strategies and random pricing strategies, the simulation results show that the price strategy based on DRL is more profitable, which further explains the effectiveness and superiority of the price strategy based on DRL.**

Keywords—Smart grid; Electric energy transaction; Reinforcement learning; Dynamic pricing

I. INTRODUCTION

In the smart grid system, when the power price in the power market rises or the reliability of the power system is threatened during the peak period, the demand response [2] can be used to reduce or delay the power load in a certain period. However, due to the lack of real-time load demand and power consumption pattern information of power users, as well as the fluctuation of power prices in the wholesale market, it is very challenging for intermediaries to actually develop effective dynamic pricing strategies [3].

Reinforcement learning has become one of the effective ways to solve complex dynamic decision-making problems with the development of artificial intelligence. Ruelens et al. proposed a modeless Monte Carlo method based on state-behavior value function and applied the fitted q iterative algorithm to the demand response environment of power users by taking household electric water heater as the research object [7]. In view of the relatively fast price fluctuation in the power market, Kim et al. studied the dynamic pricing problem in the microgrid system and used the Q-learning algorithm to develop dynamic pricing strategies for intermediaries. Wang [11] et al. proposed a hybrid learning method combining unsupervised learning, supervised learning, and reinforcement learning, and developed a smart trading strategy that can better adapt to the dynamics and complexity of the smart grid market. Yang et al. adopted the cyclic deep Q network (DRQN) method to model the power trading process and developed an effective dynamic pricing strategy by using multi-agent reinforcement learning [13].

Considering the uncertainty of the smart grid environment, this paper constructs an electric energy trading model based on intermediaries and uses a reinforcement learning algorithm to develop a dynamic pricing strategy with maximum benefits. The main work of this paper is as follows:

- We built an electric energy trading model based on an intermediary. The model includes generators, intermediaries, and electricity users. We assumed that each participant only cares about their costs and profits.
- We modeled the dynamic pricing process as the Markov decision process (MDP). Based on the model-free reinforcement learning method, a dynamic pricing strategy with unknown state transition probability was developed for intermediaries.
- To improve the convergence speed of the reinforcement learning algorithm, continuous variables such as time and price are discretized to reduce the input and output space in this paper.

The rest of this paper is organized as follows. In Section II, we briefly introduce electricity trading models and formulas. In Section III, we introduce the design of the main elements of MDP and the dynamic pricing solution based on the DQN method in detail. In Section IV, we introduce the experimental studies, and finally, we conclude our paper in Section V.

II. ELECTRIC ENERGY TRADING MODEL

The electric energy trading model constructed in this paper is shown in Fig 1. Demand response refers to the process in which power users adjust their electricity consumption according to the retail price information released by intermediaries. It is the behavior of users after considering their own electricity cost [17]. This paper assumes that all users have similar consumption pattern groups, that is, the same expected price p_C^t and the same demand response pattern. Fig 2 shows the process of the users' demand response model.

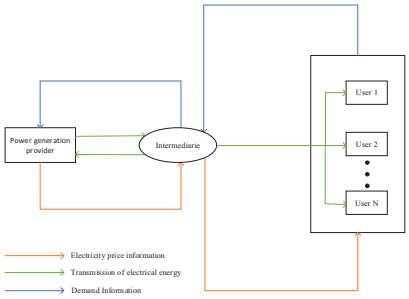

Fig 1 Electric energy trading model

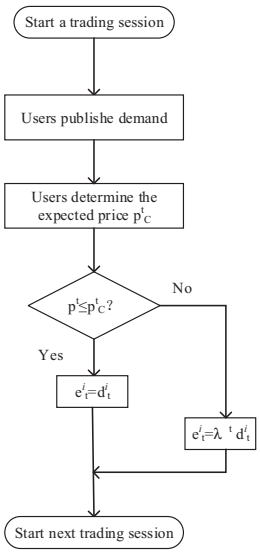

Fig 2 Users' demand response process

λ^t is the ratio of what the user actually bought to what they wanted. It's a number between 0 and 1. Its formula is

$$\lambda^t = \delta \cdot \left(\sigma - p^t\right) \tag{1}$$

Where δ and σ are the demand response factors. The demand response process can be expressed as

$$e_i^t = \begin{cases} d_i^t & if\ p^t \le p_C^t \\ \lambda^t \cdot d_i^t & otherwise \end{cases} \tag{2}$$

At time *t*, the initial electricity demand of users is d_i^t. The intermediary buys $\sum_{i\in N} d_i^t$ of electricity at the price c^t and sets the retail price p^t. The users determine the power purchase e_i^t according to the demand response. For the remaining electricity $\sum_{i\in N}\left(d_i^t - e_i^t\right)$, the intermediary needs to sell this part of the electricity back to the wholesale market at the back price b^t. At the end of time t, the cost ψ^t of buying electricity for intermediaries can be expressed as

$$\psi^t = c^t \cdot \sum_{i\in N} d_i^t - b^t \cdot \sum_{i\in N}\left(d_i^t - e_i^t\right) \tag{3}$$

The income m^t from electricity sales can be expressed as

$$m^t = p^t \cdot \sum_{i\in N} e_i^t \tag{4}$$

The total returns r^t of intermediaries in time period t can be expressed as

$$r^t = m^t - \psi^t \tag{5}$$

III. DRL Method For Dynamic Pricing

The dynamic pricing process of the intermediary can be represented by the Markov decision process as follows

$$M^B = \left(S, A, P, R, \gamma\right) \tag{6}$$

Where S is a set of states, each s_t represents the state of intermediary B at t; A is a set of actions, each a_t represents the action of intermediary B at t; $P\left(s,a\right) \to s'$ is the state transition probability function, representing the probability of the agent transitioning from state s to state s' after acting; $r \in R$ represents the immediate return of the intermediary; γ is the discount factor; $\pi\left(s\right)$ specifies the action that the intermediary should select in the state s. In this paper, the status information of the agent is

$$s_t = \left(t, p^t\right) \tag{7}$$

The action set is

$$A = \left\{Raise; Lower; Remain\right\} \tag{8}$$

Where *Raise* indicates that the intermediary raises the retail price by \$0.01; *Lower* indicates the intermediary lowered the retail price by \$0.01. *Remain* indicates that the intermediary publishes the same retail price as the previous period. The immediate return r_t is

$$r_t = p^t \cdot \sum_{i\in N} e_i^t - c^t \cdot \sum_{i\in N} d_i^t + b^t \cdot \sum_{i\in N}\left(d_i^t - e_i^t\right) \tag{9}$$

The cumulative return *R* is

$$R = \sum_{t\in T} \gamma^t r_t \tag{10}$$

Considering the advantages of DQN, this paper will use the DQN algorithm to develop the optimal dynamic pricing strategy for intermediaries. In the electricity trading model, the dynamic pricing process of the intermediary based on the DQN algorithm is shown in Table 1.

IV. Performance Evaluation

It is ideal to conduct data integrity attacks on DC lines, but in reality, it is all AC lines. Chapter II of this paper uses Kalman filter technology to estimate the system's state. Taking EKF technology as an example, in order to successfully attack data integrity, it is necessary to establish a data integrity attack model for EKF technology.

A. Parameter Setting

This paper considers a generator, an intermediary, and 200 power users. Time is discretized into a set of time periods $T = \left\{t = 1, 2, \ldots, 24\right\}$, where each session lasts one hour. The electric energy trading process was carried out continuously for 10 days in a round of training. So, there are 240 sessions in a training round and the total return of the intermediary is also the sum of 10 days' returns. It is relatively reasonable to assume that the demand for electricity per hour is between 0.2kWh and 0.8kWh. The retail electricity price issued by the intermediary is limited to \$0.08~\$0.2, and changes by 0.01 step size. The users' expected price p_C^t is set at \$0.10. The users' demand response factor is set to $\delta = 5, \sigma = 0.3$. The intermediary's expected profit for 1kwh of electricity is set as

Table 1 DQN algorithm process

Algorithm 1 DQN algorithm

Example Initialize playback memory D with capacity n;
The random weights θ are used to initialize the action-behavior function Q;
Initializing $\theta^- = \theta$ and calculating the action-behavior value of the target network;
for *episode*=1, *M* do:
  Initializing the environment to get the initial state s_1;
  for t=1, T do:
    To randomly select an action a_t, or select the action with the maximum value function
    $a_t = argmax_a Q^*(s_t, a; \theta)$
    Performing the action a_t to get the next state s_t and return r_t;
    Storing (s_t, a_t, r_t, s_{t+1}) in the playback memory D;
    A uniform random sampling (s_j, a_j, r_j, s_{j+1}) from the playback memory D;
    Calculating $y_j = \begin{cases} r_j & if\ s_{t+1}\ is\ final\ state \\ r_j + \gamma \max_{a'} Q(s_{j+1}, a'; \theta^-) & otherwise \end{cases}$;
    The gradient descent method is used to solve the problem according to $\left(y_j - Q(s_j, a_j; \theta)\right)^2$;
  Completing parameter updates $\theta^- \leftarrow \theta$ at intervals C;
  end for
end for

\$0.03. Both the wholesale price and the resale price are fixed and they are $c^t = 0.05, b^t = 0.03$. The parameter Settings of reinforcement learning are shown in Table 2.

Table 2 Reinforcement learning parameters setting

Reinforcement learning parameters	Values
Rate of learning α	0.001
Discount factors γ	0.7
Target network update frequency	10
Coefficient of exploration ε	0.2
Number of learning rounds	100

B. *Dynamic Pricing Simulation*

Fig 3 shows the results of the intermediary's income for three pricing strategies. We can see that with a fixed price, the intermediary's return is around 1200, with little fluctuation, but the return is also very low. With uniform random pricing, the intermediary's return rose to about \$1450, but the volatility also increased. Based on the dynamic pricing strategy of the DQN algorithm, the return of the intermediary is maintained at a high level, and the fluctuation is not obvious after 30 rounds of training. It indicates that the intermediary has learned an optimal dynamic pricing strategy.

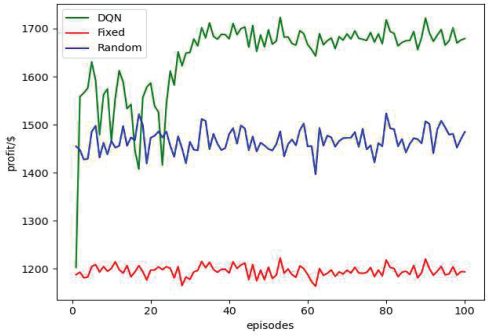

Fig 3 Intermediary's returns under different strategies

C. *Demand Response Simulation*

In the process of user-side demand response, this paper introduced the variable demand response coefficient λ^t. The calculation method is given by formula (2-1). This paper redefines the demand response coefficient as follows

$$\lambda^t = \begin{cases} 1 & if\ p^t \le p_C^t \\ \delta \cdot (\sigma - p^t) & otherwise \end{cases} \tag{11}$$

The purchase amount of the users can be expressed as

$$e_i^t = \lambda^t \cdot d_i^t \tag{12}$$

Therefore, the relationship between the demand response coefficient λ^t and the retail price p^t published by the intermediary can be obtained as shown in Fig 4. When the retail price is \$0.08~\$0.10, the demand response coefficient is always 1. It means that when the price released by the intermediary is low, the user will buy exactly according to the demand. When the retail price is \$0.10~\$0.20, the demand response coefficient is a first-order function of negative correlation with the retail price. It indicates that the electricity purchased by users decreases with the increase of the retail price.

In this paper, the retail price data released by the intermediary in 24 periods on the last day of training is used to obtain the change of the users' demand response coefficient at the time of the day. Fig 5 shows the change of the retail price and users' demand response coefficient over time. We can acknowledge that the average power purchased by users according to the demand response is only equivalent to the published demand after analysis.

Fig 6 shows the total demand and actual total power purchase for 200 users in one day. We can see that the total demand of the user fluctuates a little around 100, while the purchase of electricity fluctuates greatly. It is mainly caused by the dynamic change of the retail price published by the intermediary. The demand response of power users

can effectively limit the overpricing of the intermediary. It plays an important role in maintaining the price stability of the electricity market and the balance between the supply and demand of electricity trading.

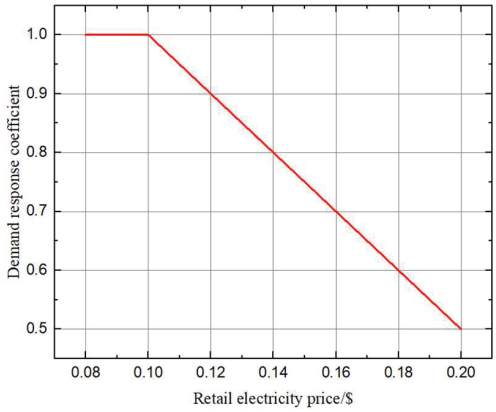

Fig 4 Relation between demand response coefficient and retail price

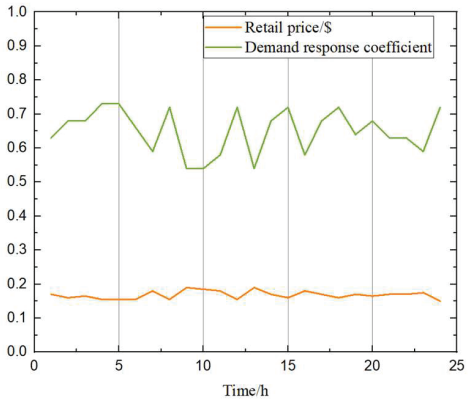

Fig 5 Demand response coefficient and retail price in one day

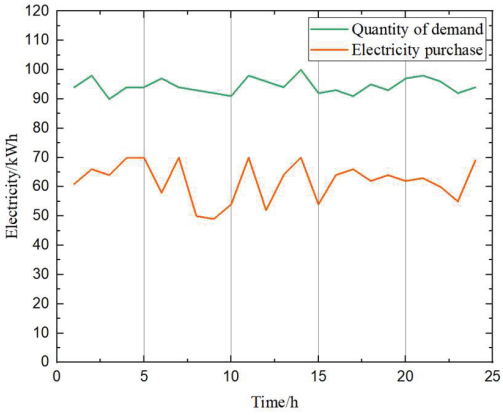

Fig 6 Total user demand and total power purchase in a day

D. *The Impact of Price Fluctuations*

To verify the stability and applicability of the dynamic pricing strategy of the intermediary based on the reinforcement learning DQN algorithm, this paper makes the wholesale price vary from $0.04 to $0.06 in the experiment. At the same time, fixed pricing strategy and random pricing strategy were introduced as a comparison. The profits of the three methods are shown in Fig 7. As can be seen from the figure, the fluctuation range of intermediary returns of the three pricing strategies is larger than before. It is directly caused by the fluctuation of wholesale prices.

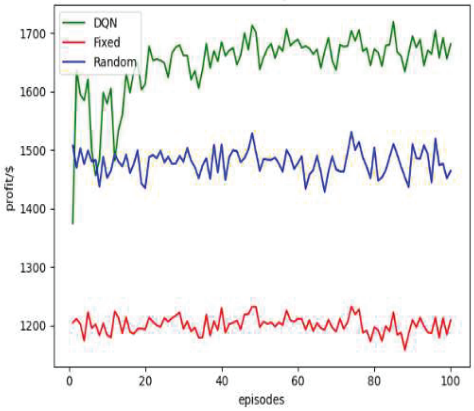

Fig 7 Returns of intermediaries when wholesale prices fluctuate

It can be seen that the benefit of the DQN method is significantly higher than that of the other two methods when the episodes are more than 20. From the point of view of the data, the intermediary's income and the proportion of increase change less than that of the other two methods. It also verifies the stability and applicability of dynamic pricing strategy based on reinforcement learning DQN algorithm.

V. Conclusion

This paper presents a dynamic pricing mechanism for the intermediary in the process of electricity trading. We introduced the reinforcement learning method and used the DQN algorithm to develop the optimal dynamic pricing strategy. Based on the discretization of continuous variables, the state space, and action space are reduced, and the convergence speed of the reinforcement learning algorithm is improved. This paper designed the experiment to simulate the electric energy trading model. The results show that the dynamic pricing strategy based on the DQN algorithm in reinforcement learning can significantly improve the returns of the intermediary. At the same time, this paper also studies the wholesale market power price fluctuations on the intermediary's income and proves the effectiveness and superiority of the DQN algorithm dynamic pricing strategy.

References

[1] Tuballa M L, Abundo M L. A review of the development of Smart Grid technologies[J]. Renewable and Sustainable Energy Reviews, 2016, 59: 710-725.

[2] Paterakis N G, Erdinç O, Catalão J P S. An overview of Demand Response: Key-elements and international experience[J]. Renewable and Sustainable Energy Reviews, 2017, 69: 871-891.

[3] Kim B G, Zhang Y, Van Der Schaar M, et al. Dynamic pricing for smart grid with reinforcement learning[C]//2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, 2014: 640-645.

[4] Sutton R S, Barto A G. Reinforcement learning: An introduction[M]. MIT press, 2018.

[5] Zhang D, Han X, Deng C. Review on the research and practice of deep learning and reinforcement learning in smart grids[J]. CSEE Journal of Power and Energy Systems, 2018, 4(3): 362-370.
[6] Chen Y, Huang S, Liu F, et al. Evaluation of reinforcement learning-based false data injection attack to automatic voltage control[J]. IEEE Transactions on Smart Grid, 2018, 10(2): 2158-2169.
[7] Ruelens F, Claessens B, Vandael S, et al. Residential Demand Response of Thermostatically Controlled Loads Using Batch Reinforcement Learning[J]. IEEE Transactions on Smart Grid, 2017, 8(5): 2149-2159.
[8] Peters M, Ketter W, Saartsechansky M, et al. A reinforcement learning approach to autonomous decision-making in smart electricity markets[J]. Machine Learning, 2013, 92(1): 5-39.
[9] Reddy P P, Veloso M M. Negotiated learning for smart grid agents: entity selection based on dynamic partially observable features[C]//Twenty-Seventh AAAI Conference on Artificial Intelligence. 2013.
[10] Kim B, Zhang Y, Der Schaar M V, et al. Dynamic Pricing and Energy Consumption Scheduling With Reinforcement Learning[J]. IEEE Transactions on Smart Grid, 2016, 7(5): 2187-2198.
[11] Wang X, Zhang M, Ren F, et al. A hybrid-learning based broker model for strategic power trading in smart grid markets[J]. Knowledge Based Systems, 2017: 142-151.
[12] Ghosh S, Subramanian E, Bhat S P, et al. VidyutVanika: A reinforcement learning based broker agent for a power trading competition[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33: 914-921.
[13] Yang Y, Hao J, Sun M, et al. Recurrent Deep Multiagent Q-Learning for Autonomous Brokers in Smart Grid[C]. international joint conference on artificial intelligence, 2018: 569-575.
[14] Urieli D, Stone P. Autonomous electricity trading using time-of-use tariffs in a competitive market[C]//Thirtieth AAAI Conference on Artificial Intelligence. 2016.
[15] Chen T, Su W. Indirect customer-to-customer energy trading with reinforcement learning[J]. IEEE Transactions on Smart Grid, 2018, 10(4): 4338-4348.
[16] Ketter W, Collins J, Saar-Tsechansky M, et al. Information systems for a smart electricity grid: Emerging challenges and opportunities[J]. ACM Transactions on Management Information Systems (TMIS), 2018, 9(3): 1-22.
[17] Le D T, Zhang M, Ren F. An economic model-based matching approach between buyers and sellers through a broker in an open e-marketplace[J]. Journal of Systems Science and Systems Engineering, 2018, 27(2): 156-179.
[18] González A Y R, Alonso M P, Lezama F, et al. A competitive and profitable multi-agent autonomous broker for energy markets[J]. Sustainable Cities and Society, 2019, 49: 101590.
[19] Hansen T M, Chong E K P, Suryanarayanan S, et al. A partially observable markov decision process approach to residential home energy management[J]. IEEE Transactions on Smart Grid, 2016, 9(2): 1271-1281.
[20] Szepesvári C. Algorithms for reinforcement learning[J]. Synthesis lectures on artificial intelligence and machine learning, 2010, 4(1): 1-103.
[21] Arulkumaran K, Deisenroth M P, Brundage M, et al. Deep reinforcement learning: A brief survey[J]. IEEE Signal Processing Magazine, 2017, 34(6): 26-38.
[22] Ge H, Song Y, Wu C, et al. Cooperative deep Q-learning with Q-value transfer for multi-intersection signal control[J]. IEEE Access, 2019, 7: 40797-40809.
[23] Tokic M, Palm G. Value-difference based exploration: adaptive control between epsilon-greedy and softmax[C]//Annual Conference on Artificial Intelligence. Springer, Berlin, Heidelberg, 2011: 335-346.
[24] Henderson P, Islam R, Bachman P, et al. Deep reinforcement learning that matters[C]//Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
[25] François-Lavet V, Henderson P, Islam R, et al. An introduction to deep reinforcement learning[J]. arXiv preprint arXiv:1811.12560, 2018.
[26] Hester T, Vecerik M, Pietquin O, et al. Deep q-learning from demonstrations[C]//Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
[27] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.
[28] Özdemir S, Unland R. AgentUDE17: A genetic algorithm to optimize the parameters of an electricity tariff in a smart grid environment[C]//International Conference on Practical Applications of Agents and Multi-Agent Systems. Springer, Cham, 2018: 224-23.