Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Optimal Electricity Decomposition Method for New Energy Grid-Connected System based on Q-Learning Algorithm

To cite this article: Jinhui Chen *et al* 2022 *J. Phys.: Conf. Ser.* **2310** 012033

Optimal Electricity Decomposition Method for New Energy Grid-Connected System based on Q-Learning Algorithm

Jinhui Chen[1a*], Zi Yang[2a*], Zhengfeng Wang[2b*], Na Yang[3a*], Rong Fu[1b*]

[1] Nanjing University of Posts and Telecommunications, Nanjing210023, China

[2] State Grid Anhui Electric Power Co.,Ltd., Hefei 230022, China

[3] Institute of Economy and Technology, State Grid Anhui Electric Power Co., Ltd., Hefei 230022, China

[1a*]17302559635@163.com, [2a*]hit_yangzi@163.com, [2b*]2425825662@qq.com, [3a*]nyang_sjtu@163.com, [1b*]399010550@qq.com

Abstract—Since the traditional contract power decomposition in the power trading market is difficult to meet the needs of new energy participating in the system operation, an optimal decomposition method of contract power based on the Q-learning algorithm under the uncertainty of new energy power is proposed. Considering the uncertainty of new energy power output, an optimization model to minimize the power purchase cost of the power grid is established. Given the uncertainty of the newly added electricity and the quotation in the market transaction, it is proposed to use the enhanced Q-learning algorithm to obtain the contract decomposition of electricity. According to the actual annual contract electricity data, the monthly optimal decomposition results of contract electricity are obtained, which verifies the economy and effectiveness of the optimal electricity decomposition method.

1. Introduction

The decomposition of contract electricity is completed by multiple time scales successively. The traditional contract electricity is decomposed by monthly peak and valley electricity, or by unit electricity. In [1], the electricity purchaser aims to minimize the cost of electricity purchase, and the target electricity is decomposed into contract electricity and new electricity. In [2], in the process of decomposing the annual contract electricity of the power station, the influence of the unit is ignored. It establishes a multi-objective function, which considered the decomposition of contract electricity from many aspects[3]. In [4], the influence of the unit in the electricity decomposition is considered, but the in-depth analysis is lacking.

Reinforcement learning achieves specific goals by facilitating the interaction between the agent and the environment. Given the above problems, we need to explore and learn the optimal trading strategy under the maximization of cumulative rewards, without the need for complex optimization calculations. Likewise, there is no need to invest time and effort into the transaction process [5].

At present, the application of reinforcement learning in the power system mainly covers many aspects such as demand response management, operation control, and economic dispatch [6]. In the field of the electricity market, it proposed an intelligent quotation strategy based on an adaptive reinforcement learning model considering factors such as market demand and historical transaction conditions[7]. In [8], reinforcement learning methods are used to implement indirect client-to-client transactions.

Given the uncertainty of photovoltaic power generation, this paper fully considers the constraints of contract peak and valley power and photovoltaic power. Under the premise that the lowest electricity purchase cost of power grid enterprises is considered as the objective function, the optimal decomposition model of annual contract electricity is established, and Q-learning algorithm is used to solve the model. Finally, the results are verified in terms of electricity purchase cost and algorithm through an example.

2. Electricity Decomposition Model Based on Chance Constrained Programming

2.1. Objective Function Based on Chance Constraint Programming

Since the monthly transaction volume and monthly load forecast of the electricity market are carried out according to the peak-valley period respectively, the peak-valley method is also used by this model to decompose the contracted electricity. In the model, the effects of new electricity and new electricity prices are considered, and the sources of new electricity are thermal power and photovoltaics. In the model, the contract peak electricity and valley electricity are used as the solution targets, and all new electricity and new electricity prices are predicted. In new energy grid-connect system, in order to minimize the electricity purchase cost of power grid enterprises, the following model can be established:

$$f = \min[\sum_{k=1}^{K}(p_{h,c,k}q_{c,h,k} + p_{l,c,k}q_{c,l,k}) + \sum_{k=1}^{K}(p_{h,f,k}q_{f,h,k} + p_{l,f,k}q_{f,l,k}) + \sum_{k=1}^{K} p_{p,k}(q_{p,h,k} + q_{p,l,k})] \quad (1)$$

where $p_{h,c,k}$ and $p_{l,c,k}$ are the contract electricity prices during peak and valley periods; $p_{h,f,k}$ and $p_{l,f,k}$ are the thermal power prices in the newly added electricity during peak and valley periods; $p_{p,k}$ is the photovoltaic electricity price in the newly added electricity; $q_{c,h,k}$, $q_{f,h,k}$ and $q_{p,h,k}$ are the contracted electricity during the peak period, the thermal power among the newly added electricity, and the photovoltaic electricity among the newly added electricity; $q_{c,l,k}$, $q_{f,l,k}$, and $q_{p,l,k}$ are the contracted electricity during the valley period, the thermal power in the newly added electricity, and the photovoltaic electricity in the newly added electricity; K is the total number of months.

2.2. Restrictions

(1) Unit power constraints

The unit power constraint can be described as:

$$q_{c,h,k} + q_{f,h,k} + q_{p,h,k} = q_{h,k} \quad (2)$$

$$q_{c,l,k} + q_{f,l,k} + q_{p,l,k} = q_{l,k} \quad (3)$$

$$\sum_{k=1}^{K}(q_{c,h,k} + q_{c,l,k}) = Q_c^{tol} \quad (4)$$

$$q_{c,h,k} + q_{c,l,k} = Q_c^{t} \quad (5)$$

where $q_{h,k}$ and $q_{l,k}$ are the total electricity during the peak period and the total electricity during the valley period in the month k , respectively; Q_c^{tol} is the total electricity in the annual contract; Q_c^t is the contracted electricity in the month k .

(2) Power upper and lower limit constraints

$$lq_{h,k} \le q_{c,h,k} \le iq_{h,k} \quad (6)$$

$$uq_{l,k} \le q_{c,l,k} \le vq_{l,k} \quad (7)$$

where l and i are the minimum and maximum parameters of the contracted electricity during the peak periods in the month k , respectively. u and v are the minimum and maximum parameters of the contracted electricity during the valley periods.

3. Q-learning Algorithm Solution

3.1. The state space of the system

The system state vector s_k consists of the following quantities: decision moment t_k, predicted value of thermal power peak electricity price in newly added electricity $p^y_{h,f,k}$, predicted value of thermal power trough electricity price in newly added electricity $p^y_{l,f,k}$, monthly peak total electricity forecast value $q^y_{h,k}$ and monthly trough total electricity forecast value $q^y_{l,k}$ .Specifically, as shown in formula (8):

$$s_k = (t_k, p^y_{h,f,k}, p^y_{l,f,k}, q^y_{h,k}, q^y_{l,k}) \in \Phi \tag{8}$$

where Φ is the system state space, which represents the set of all state vectors.

3.2. The action vector and the reward function of the system

Action refers to the reflection of each agent in response to environmental changes. The various responses form an action set similar to (a_1, a_2, a_3). In this paper, a similar action set is given in the peak power and the valley power respectively. In order to make the results of Q-learning more accurate, 0.001TWh is taken as Δ value. The action vectors of the peak contract electricity and the trough contract electricity to be sought are $\{0, \Delta, 2\Delta, 3\Delta, \cdots, q_{c,h,k} + q_{c,l,k}\}$ and $\{0, \Delta, 2\Delta, 3\Delta, \cdots, q_{c,h,k} + q_{c,l,k}\}$, respectively.

Define system policy π as a set of state space-action mappings. Under this strategy, the system selects action a_k under the current state s_k and constraints, and transfers to the next decision-making period through strategy π. The peak-shaving compensation cost r_k is:

$$r_k = p_{h,c,k} q_{c,h,k} + p_{l,c,k} q_{c,l,k} + p_{h,f,k} q_{f,h,k} + p_{l,f,k} q_{f,l,k} + p_{p,k}(q_{p,h,k} + q_{p,l,k}) \tag{9}$$

The action a_k system is selected according to the strategy π in the current period, and the optimal expected value $Q^\pi(s_k, a_k)$ is also obtained. Based on the stochastic dynamic programming method, the system optimization objective is transformed into selecting an optimal strategy $\pi^*(s_k)$ from the set of strategies for generating action a_k from the system state s_k, so as to minimize the cost r_k:

$$\pi^*(s_k) = \arg\min_{\pi\in\Omega} Q^\pi(s_k, a_k) = \arg\min_{\pi\in\Omega} E_\pi \left| \eta \sum_{k=0}^{K} r_k(s_k, a_k) \right| \tag{10}$$

where $Q^\pi(s_k, a_k)$ is the expected operating cost of the system in period k ; Ω is the set of all strategies π ; η is the discount factor, and $0 < \eta < 1$.

3.3. The specific solution process

The specific solution process of the Q-learning algorithm is as follows, and the process is shown in Figure 1:

Step 1: Initialize the Q-value table. Input total number of learning sample tracks M, sample orbit decision period number K, learning rate α, learning rate update coefficient μ_α, discount factor η. Let sample track $m = 0$.

Step 2: Let $k = 0$. Initialize system state data such as decision moment, predicted value of thermal power peak electricity price in newly added electricity, predicted value of thermal power trough electricity price in newly added electricity, monthly peak total electricity forecast value, monthly trough total electricity forecast value.

Step 3: The greedy strategy is $\pi(a_k) = \begin{cases} a_{random}, random(0,1) \le \varepsilon \\ a_{greedy}, random(0,1) > \varepsilon \end{cases}$. If the randomly generated number $random(0,1)$ is less than the greedy exploration probability ε, select the random action a_{random} and execute it. then randomly get the next state; otherwise, choose action a_k based on the Q value: $a_{greedy} = \arg\min Q(s_k, a_k)$.

Step 4: Calculate the system cost r_k generated by executing action a_k in state s_k during decision period k . Update the Q value and update the strategy at the same time:

$Q^*(s_k,a_k)=(1-\alpha)Q(s_k,a_k)+\alpha[r_k+\min Q(s_{k+1},a_{k+1})]$. Continuously learn and update the minimum Q value in the state-action set, and use the corresponding peak-valley contract power plan as the latest action to achieve the purpose of optimizing the strategy. Let $k:=k+1$ and go back to step 3; If $k=K+1$, go to step 5.

Step 5: Execute the action a_k selected by the current state s_k, and calculate the cost r_k and the final state cost $r(s_{k+1})$ generated in the state transition process. Update the Q value table according to formula (11), and update the strategy at the same time. Let $m:=m+1$, $\alpha:=\mu_\alpha*\alpha$.

$$Q^*(s_k,a_k)=(1-\alpha)Q(s_k,a_k)+\alpha\left[r_k+r(s_{k+1})\right] \tag{11}$$

Step 6: If $m<M$, go back to step 2; Otherwise, the program ends and outputs the obtained contracted electricity during the peak period and contracted electricity during the valley period.

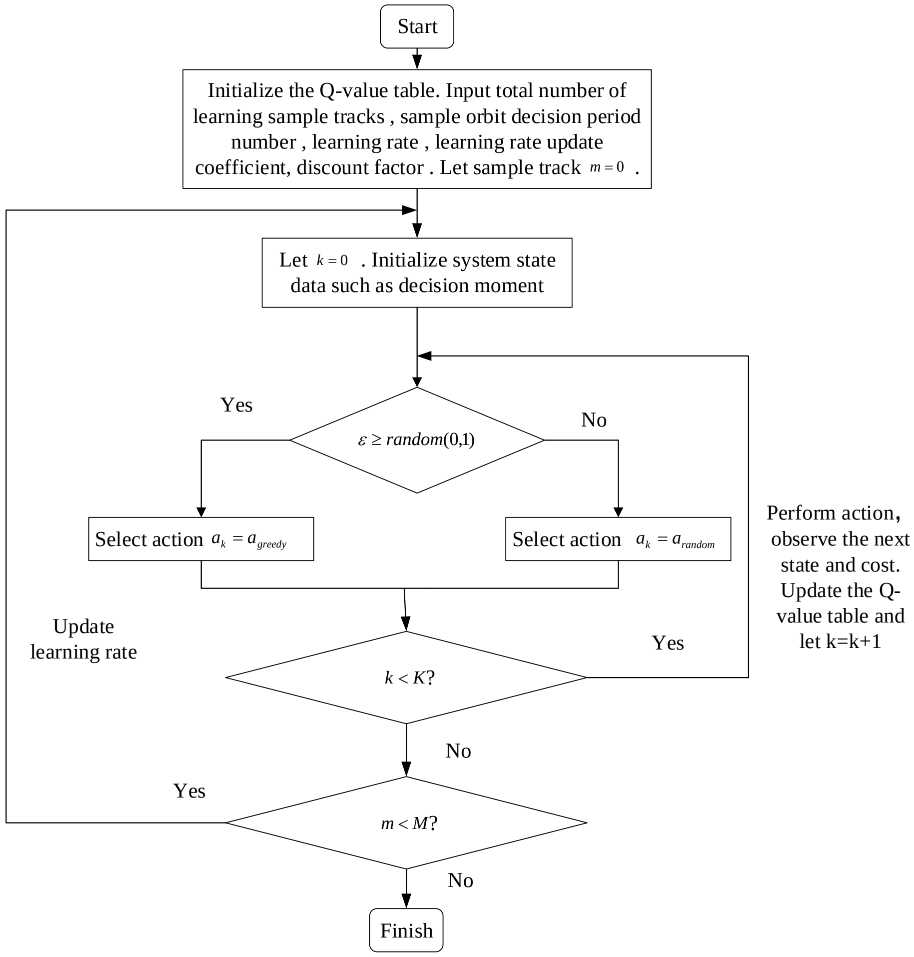

Fig.1 Q learning algorithm solution steps

4. Case analysis

Taking a medium and long-term contract transaction as an example, the total contract electricity is 10TWh, and the contract execution time is one year. In order to simplify the calculation model, the peak-to-valley electricity decomposition method is adopted. 22:00-8:00 the next day is the trough period. The contract peak electricity price is 248 yuan/MWh, and the contract trough electricity price is 118yuan/MWh. The generator units are three groups of 600MW thermal power units and one group of photovoltaic units. The electricity in the contract and the newly added electricity each month are generated by the thermal power unit, and the photovoltaic power is generated by the photovoltaic generator unit. In the given optimization model, l is 0.4, i is 0.9, u is 0.05, v is 0.3, and the time interval k is one month.

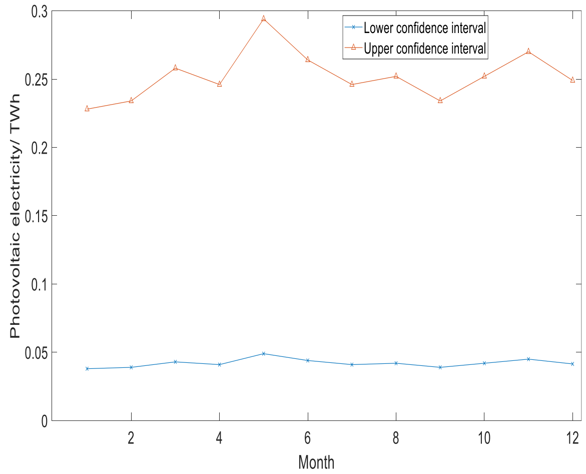

Fig.2 Confidence interval for the photovoltaic electricity among the newly added electricity

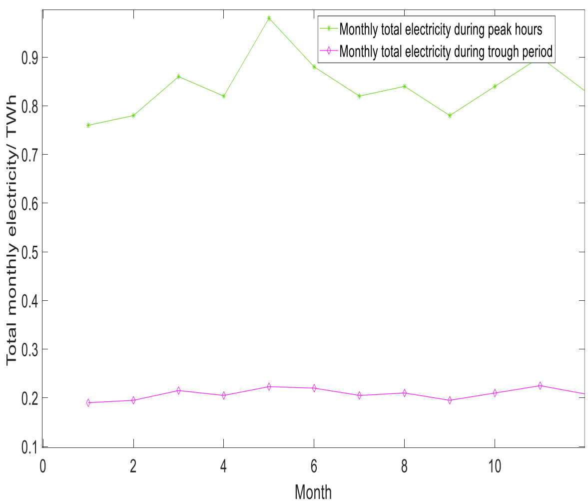

Fig.3 Predicted monthly total power

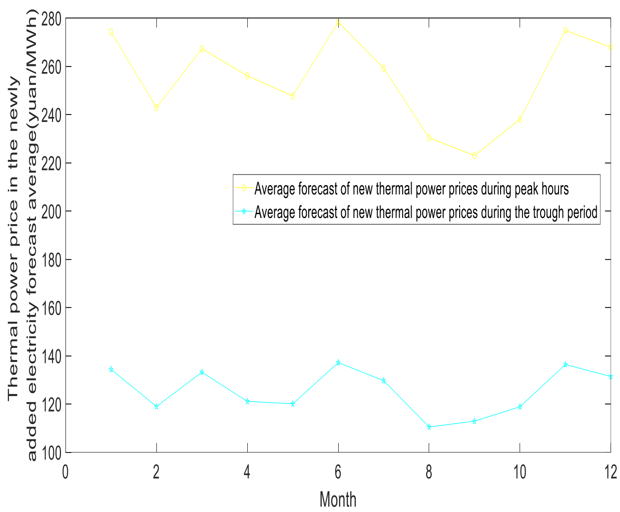

Fig.4 Thermal power price in the newly added electricity forecast average

Since the power grid and power generation companies obtain the clearing electricity price through the separate bidding and clearing during the peak and valley periods each month, load forecasting and electricity price forecasting should also be carried out according to the peak and valley periods respectively. The forecast of the photovoltaic electricity in the newly added electricity confidence interval is shown in Figure 2. The forecast of the total electricity during the peak and valley period is shown in Figure 3. The forecast of the average forecast of thermal power price in newly added electricity is shown in Figure 4.

Considering the fluctuation of electricity prices and many uncertain factors in electricity market transactions, we use the Q-learning algorithm to solve the monthly electricity quantity. In order to take into account the accuracy and uncertainty, we take three groups of the obtained solutions to obtain the average value, and the results are shown in Table 1. Then we calculate the total electricity purchase cost obtained by different methods, and the result is shown in Figure 5.

As shown in Figure 5, the total electricity purchase cost obtained by solving the model in this paper according to the Q-learning algorithm is 2.793 billion yuan, and the total electricity purchase cost obtained by using the nonlinear algorithm to solve the model in this paper is 2.799 billion yuan. The cost of electricity purchase is 2.812 billion and 2.821 billion yuan respectively. Instead of using the model in this paper, using the decomposition method that only decomposes the monthly electricity according to the proportion will cost 2.889 billion yuan to purchase electricity. Obviously, the method of Q-learning is used to solve the problem, which allows us to obtain better economic benefits while fully considering the uncertainty of electricity price and electricity.

Tab.1 Q-learning algorithm breaks down results by monthly electricity

Month	Power breakdown results during peak hours /TWh	Decomposition results of electricity during the trough period /TWh
1	0.684	0.112
2	0.469	0.165
3	0.774	0.16
4	0.728	0.139
5	0.725	0.2
6	0.792	0.144
7	0.695	0.181
8	0.492	0.189
9	0.478	0.176
10	0.667	0.189
11	0.806	0.149
12	0.747	0.149

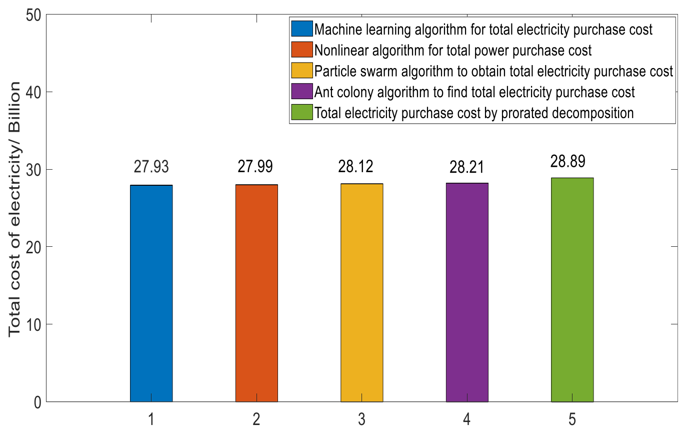

Fig.5 Cost comparison of three power decomposition methods

Not only that, as a widely used reinforcement learning algorithm, Q-learning has many advantages. Since Q-learning is based on temporal difference solution, its principle is simple, the required parameters are few, the acceptable range is wide, and the evaluation strategy is clear. It has better resource occupancy and solution speed in dealing with the uncertain problem in the calculation example. Because Q-learning has better timeliness and robustness, the prediction of photovoltaic output value is more accurate. Therefore, it can well meet the online real-time optimization requirements of the power system.

5. Conclusion

This paper proposes the optimal decomposition model of contract electricity under the uncertainty of new energy, and solves the contracted electricity during the peak and the valley period. Fully considering the uncertainty of photovoltaic output, the Q-learning algorithm is used to efficiently complete the solution. Through the example analysis, the following conclusions can be drawn:

(1) Compared with the traditional contract power decomposition method, the contract power decomposition model proposed in this paper fully takes into account the uncertainty and randomness of photovoltaic power, the impact of thermal power unit output, electricity price forecasting and load forecasting, making the decomposition results more convincing.

(2) When solving, this paper adopts the Q-learning algorithm to solve. When the peak and valley electricity is decomposed in the model, each different decomposition method will have different electricity purchase costs, which is also in line with the characteristics of Q-learning trial and error and delayed reward. In the case of learning a better learning strategy based on this paper, the Q-learning algorithm can get results faster and more efficiently than the ordinary algorithm.

Acknowledgments

This work was financially supported by the State Grid Anhui Electric Power Company Technology Pr oject No. SGAHJY00ZLJS2200049 (Research on key mechanism of power spot market operation and capacity compensation for "dual carbon" target).

References

[1] Zhang Shaodi. The decomposition method of annual contract elec tricity based on CSS[J]. Electric Power Automation Equipment, 2014, 34(11): 135-141.

[2] Huang Qiang, Sun Xiaoyi, Zhang Hongbo, Hao Peng. Determination and Decomposition of Contract Electricity for Cascade Hydropower Stations on the Upper Han River[J]. Journal of Hydroelectric Engineering,2011,30(04):246-252.

[3] Miao Shumin, Luo Bin, Shen Jianjian, Cheng Chuntian, Li Gang, Sun Yongjun. Hydro-thermal power short-term multi-objective power generation dispatch considering market transition and medium- and long-term contract power decomposition[J].Power System Technology,2018,42(07):2221-2231 .

[4] Dong Li, Gao Ciwei, Yu Jie, Teng Xianliang, Tu Mengfu, Ding Qia. Frequency modulation backup market mechanism considering the decomposition of medium and long-term electricity contracts[J].Automation of Electric Power Systems,2018,42(14):61-66+74.

[5] Han Dong, Huang Wei, Yan Zheng. Virtual bidding strategy for power market based on deep reinforcement learning[J/OL].Proceedings of the CSEE,{3},{4}{5}:1-14[2021-07-19] .

[6] Liu Guojing, Han Xueshan, Wang Shang, Yang Ming, Wang Mingqiang. Wind-storage cooperative decision-making based on reinforcement learning method[J].Power System Technology,2016,40(09):2729-2736.

[7] Y. Hou, Y. Ong, L. Feng and J. M. Zurada, "An Evolutionary Transfer Reinforcement Learning Framework for Multiagent Systems," in IEEE Transactions on Evolutionary Computation, vol. 21, no. 4, pp. 601-615, Aug. 2017.

[8] E. Chalmers, E. B. Contreras, B. Robertson, A. Luczak and A. Gruber, "Learning to Predict Consequences as a Method of Knowledge Transfer in Reinforcement Learning," in IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 6, pp. 2259-2270, June 2018.