RESEARCH ON DISTRIBUTED RENEWABLE ENERGY TRANSACTION DECISION-MAKING BASED ON MULTI-AGENT BILEVEL COOPERATIVE REINFORCEMENT LEARNING

Zhangyu CHEN
Shanghai Jiaotong University – China
Los_oday@sjtu.edu.cn

Dong LIU
Shanghai Jiaotong University – China
Dongliu@sjtu.edu.cn

Xiaofei Wu, Xiaochun Xu
State Grid Huai'an Power Supply Company
sease@163.com

ABSTRACT

With more and more distributed renewable energy connecting to the distribution network, the issue of regional transaction of distributed renewable energy has attracted more and more attention. In order to adapt to the complex transaction decision-making problem, this paper proposes a multi-agent bilevel cooperative reinforcement learning algorithm under the framework of bilevel stochastic decision-making model. By constructing a bilevel stochastic decision-making optimization model for distributed renewable energy trading, the uncertainties and fluctuations of distributed generation output are effectively solved. The objective of upper level planning is to maximize the profits of distributed renewable energy generators. The lower level planning is to optimize the dispatch of the whole regional market. The two layers are continuously iterated until the lower level planning is optimal, that is, the comprehensive benefit is maximized. After introducing multi-agent bilevel cooperative reinforcement learning, the algorithm can effectively carry out learning training, and after completing the training, it can quickly and accurately calculate the optimal results. Through the simulation of the model project of Guizhou Hongfeng area, the bidding decision algorithm has been verified, which can improve the profit of the power producer while taking risks into consideration, and at the same time maximize the comprehensive benefits.[1]

1 INTRODUCTION

With the progress and development of society, the global demand for green, clean and efficient electricity is increasing, so more and more distributed renewable energy is connected to distribution network.

Although there is no fuel cost for distributed photovoltaic and wind power generation, its construction cost, operation cost, and maintenance cost are high. At present, Chinese new energy distributed generators mainly make profits through national and local government electricity price subsidies. However, with the increase of distributed power penetration rate, the profit model obviously does not conform to the market rules. Subsidies to distributed generators through user subscription fees can help generators to participate in market competition with reasonable quotations based on their own potential benefits and generation costs, thus maximizing social benefits.

At present, although there are many studies on bidding strategies of power producers, there are few studies on distributed power generation in bidding, especially new energy distributed generation, considering the participation of new energy. Some studies have integrated distributed generation into the management of micro-grid or virtual power plant to participate in bidding. For example, Yu (2014) constructs a bilevel stochastic bidding planning for micro-grid considering the uncertainty of distributed generation output. Shi (2014) constructs a robust stochastic bidding model for virtual power plant considering the uncertainty of power price and the uncertainty of distributed generation output. Palma-Behnke (2005) and Li (2007) regard distributed power as a power distribution company that can control itself and incorporate it into the power purchase company's power purchase strategy. However, in a word, the new energy distributed power producers participate in the retail market, and the transaction decision optimization under the interactive transaction mode with the load side users remains to be studied.

2 BILEVEL STOCHASTIC DECISION-MAKING OPTIMIZATION MODEL FOR DISTRIBUTED RENEWABLE ENERGY TRADING

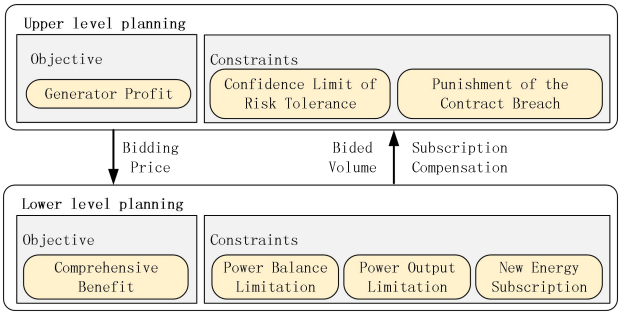

Fig.1 Bilevel stochastic decision making model for power suppliers bidding

In this paper, a bilevel stochastic programming model is

1 Manuscript received: January,14,2019. This work was supported by Science and Technology Project of State Grid of China (Research on VPP Based Source-network Coordinated Energy Management Technology of Multi-layer Distribution Network with High Proportion of Photovoltaic Generation (J20170124).)

used to solve the bidding decision-making optimization problem of distributed generators considering uncertainties. Among them, the upper level is planned to be the generator level, aiming at maximizing the income of the target distributed renewable energy generators, and the quotation of each time period is the optimization variable. When calculating the revenue, the winning bidding power and the possible new energy subscription subsidy should be given by the lower level planning. The lower level plan is optimized for the balanced scheduling of the entire regional market, and its input comes from the upper level plan, which is the quotation of each power producer. Based on the principle of the lowest operating cost of the market and satisfying the users' subscription requirements as much as possible, the optimization goal of comprehensive benefits is formulated. Considering the limitations of system power balance and user-side new energy subscription, the power generation scheduling plan of each power producer is optimized. Because the target bidder's bid information is not known to the target new energy generator, it becomes a random variable in the optimization model. Therefore, the upper optimization model will exist in the form of stochastic programming of chance constrained programming, and the target is the expected profit under a certain confidence level. At the same time, the default penalty that may be caused by the uncertainty of the distributed power supply will also be embedded in the upper-level plan in the form of expected value constraints. The two layers are continuously iterated until the upper layer plan is optimal, that is, the profit of the power producer is maximized. At this time, the bid price is the optimal decision.

2.1 Upper level planning

For the upper level planning, the model will be constructed as an opportunity constrained plan because of the uncertainty of the competitor's bidding strategy and the volatility of the distributed power supply. The objective of the optimization is to maximize the optimal value of the objective function in the form of maximax, which is a chance-constrained programming with a certain degree of confidence.

$$\begin{cases} \max \bar{f} \\ s.t. \quad f(\lambda,\xi,\varsigma)=\sum_{t=1}^{24}[\lambda^t q^{t,\xi}+c_s^{\xi}q^{t,\xi}-c_{base}q^{t,\xi}] \\ \qquad -E[W(q^{\xi},\varsigma)] \\ \qquad \Pr\{f(\lambda,\xi,\varsigma)\ge \bar{f}\} \qquad (1) \\ \qquad W(q^{\xi},\varsigma)=\gamma\sum_{t=1}^{24}q_{ub}^{t,\xi,\varsigma} \\ \qquad q_{ub}^{t,\xi,\varsigma}=\begin{cases} q^{t,\xi}-p^{t,\varsigma}T & q^{t,\xi}>p^{t,\varsigma}T \\ 0 & q^{t,\xi}\le p^{t,\varsigma}T \end{cases} \end{cases}$$

In the formula, λ represents the time-share price of the power supplier, λ^t represents the quotation for t period (decision variable), ξ represents random variables caused by unknown quotations from other bidders (random scenario simulation), ς represents the random variables (random scene simulation) caused by the uncertainty of the deviation between the real and predicted values of wind power and PVs. $f(\lambda,\xi,\varsigma)$ represents the earnings of generators in scenarios ξ and ς when quoted as λ. β represents confidence in risk taking. $\bar{f}$ represents expected returns (optimization objectives) that satisfy confidence β. $q^{t,\xi}$ represents the power supplier's standard power output in t-period (lower decision output) obtained by lower decision-making in scenario ξ. c_s^{ξ} represents the new energy subscription compensation (decision output) for unit electricity users obtained by lower level decision-making in scenario ξ. c_{base} represents unit generation cost. $W(q^{\xi},\varsigma)$ represents the default penalty of the generator in scenarios ξ and ς. γ represents a penalty for the unit's incomplete electricity, $q_{ub}^{t,\xi,\varsigma}$ represents the unbalanced power output when the standard power in scene ξ exceeds the maximum power in scene ς at time t. $p^{t,\varsigma}$ represents the upper limit of the actual output of distributed generation at t-time in scenario (random scenario simulation), T represents the length of an offer period, defaulting to 1 hour.

2.2 Lower level planning

The lower-level planning is used to target a certain bidding scenario, aiming at the comprehensive benefits of market operation, optimizing dispatching, and allocating the power of each power producer.

$$\min f=\sum_{t=1}^{24}(c_{grid}^t q_{grid}^t+\sum_{i=1}^{N_{pv}}c_{pvi}^t q_{pvi}^t+\sum_{i=1}^{N_{wp}}c_{wpi}^t q_{wpi}^t)$$
$$-\sum_{i=1}^{L}comp_{pv}Q_{loadpvi}-\sum_{i=1}^{L}comp_{wp}Q_{loadwpi}$$

$$s.t. \quad q_{grid}^t+\sum_{i=1}^{N_{pv}}q_{pvi}^t+\sum_{i=1}^{N_{wp}}q_{wpi}^t-\sum_{i=1}^{L}q_{LDi}^t=0$$
$$q_{grid}^t\ge 0, 0\le q_{pvi}^t\le q_{\max pvi}^t, 0\le q_{wpi}^t\le q_{\max wpi}^t$$
$$Q_{pv}=\sum_{t=1}^{24}\sum_{i=1}^{N_{pv}}q_{pvi}^t, Q_{wp}=\sum_{t=1}^{24}\sum_{i=1}^{N_{wp}}q_{wpi}^t, Q_{grid}=\sum_{t=1}^{24}q_{grid}^t$$
$$\upsilon_{pv}=\frac{Q_{pv}}{Q_{pv}+Q_{wp}+Q_{grid}}, \upsilon_{wp}=\frac{Q_{wp}}{Q_{pv}+Q_{wp}+Q_{grid}} \qquad (2)$$

$$Q_{loadpvi}=\begin{cases} \alpha_i\sum_{i=1}^{L}q_{LDi}^t, \upsilon_{pv}\ge\alpha_i \\ \upsilon_{pv}\sum_{i=1}^{L}q_{LDi}^t, \upsilon_{pv}<\alpha_i \end{cases}$$

$$Q_{loadwpi}=\begin{cases} \beta_i\sum_{i=1}^{L}q_{LDi}^t, \upsilon_{pv}\ge\beta_i \\ \upsilon_{pv}\sum_{i=1}^{L}q_{LDi}^t, \upsilon_{pv}<\beta_i \end{cases}$$

The lower-level planning model is actually the market-balanced scheduling model of the regional retail market. The model assigns the bidding power to the bidders.

In the formula, N_{pv} , N_{wp} and L represents the total number of PV generators in the region, the total number of wind turbines, and the total number of power users. c_{grid}^{t} indicates the cost of purchasing electricity from the external grid at time t , q_{pvi}^{t} indicates the cost of electricity purchased from PV generator i at time t (upper level planning input), c_{wpi}^{t} indicates the cost of electricity purchased from wind farmer i at time t (upper level planning input), q_{grid}^{t} indicates that electricity is purchased from the external grid at time t. q_{pvi}^{t} represents the purchase of electricity from the PV generator of i at time t (decision variable), q_{wpi}^{t} represents the purchase of electricity from the wind farmer of i at time t (decision variable), q_{LDi}^{t} indicates the load of the power user i at time t. $comp_{pv}/comp_{wp}$ indicates that the user's subscription scope is compensation for each electricity purchase in the renewable energy such as PV/wind power. q_{LDi}^{t}/q_{LDi}^{t} indicates the PV/wind power subscription power that user i should pay for on the same day. Q_{pv} , Q_{wp} and Q_{grid} indicates the amount of PV, wind, and external power consumed in the area on that day. υ_{pv} and υ_{wp} indicates the proportion of PV power generation and the proportion of wind power generation in the region, α_i and β_i Indicates the proportion of PV and the proportion of wind power subscribed by the user i. $q_{\max pvi}^{t}$ and $q_{\max wpi}^{t}$ represents the maximum power generation (upper-level planning input) at time t declared by the PV generators i and wind power generators i.

3 PRINCIPLE OF MULTI-AGENT DOUBLE-LEVEL COOPERATIVE REINFORCEMENT LEARNING ALGORITHMS

For the bilevel stochastic trading decision-making optimization model established in this paper, the conventional algorithm has been difficult to solve this model. This paper proposes a multi-agent bilevel cooperative reinforcement learning algorithm, which solves the problem of solving the bilevel stochastic decision-making model.

3.1 Single Agent Q Learning Algorithms

In the single agent Q learning algorithm, each step of the agent can select a certain action in the set of effective actions, and the environment changes under the influence of the action and gives an evaluation. The task faced by the agent is to determine an optimal strategy that maximizes the expectation of the total reward.

As shown, Q learning continually interacts with the external environment to perform value function updates on its various state-action pairs. Among them, the state-action versus value function matrix Q iteration formula is as follows:

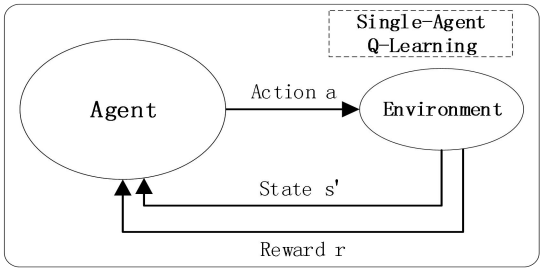

Fig.2 Schematic diagram of the single-agent Q-learning

$$Q(s,a) = (1-\alpha)Q(s,a) + \alpha[r + \gamma \max_{a' \in A} Q(s',a')] \quad (3)$$

In the formula, s, s' represent the current state and the state of the next moment, respectively. a, a' is the current action and the action at the next moment, and r is the currently awarded reward. 0< α <1, is a learning factor. 0< γ <1, is the discount factor.

3.2 Multi-Agent Bilevel Cooperative Reinforce-ment Learning Algorithms

Multi-agent double-level cooperative reinforcement learning algorithm is based on single agent Q-learning algorithm, and it is developed to solve two major problems of bilevel stochastic decision-making model. As can be seen from the figure, the outer agent I first interacts with the inner layer, outputs action a to the inner layer, and uses the reward r and state s returned from the feedback to train the agent I repeatedly, so as to update the internal control strategy of the agent I. The inner agent II interacts with the upper and lower levels of planning, parses the action a acquired by the outer layer into action A1 and a2, and inputs the two actions into the upper and lower levels of planning. The feedback states s1 and s2, reward r1 and r2, are used to train agent II to update the internal control strategy of agent II. The inner states s1 and s2, reward r1 and r2, aggregate into outer states s and reward r.

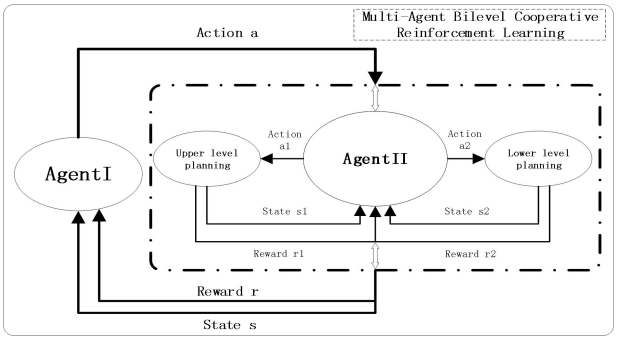

Fig.3 Multi-Agent Bilevel Cooperative Reinforce-ment Learning

By comparing the figure 1 with the figure 3, it can be found that the Agent II in the dotted line frame in the figure x becomes the intermediate link of the upper and lower layer planning. The data that needs to be exchanged between the two is used for training. The action a1 represents the winning power and subscription compensation in the upper layer plan of the figure x, and the action a2 represents the bid price of the upper layer to the lower level. Because the upper and lower layers plan to influence each other, there is a strong coupling relationship and strong randomness. The role of Agent II is to enable Agent II to obtain an optimal control strategy through repeated training, which helps the planning and calculation results of the upper and lower layers to converge quickly.

For the outer agent, the inside of the dotted line frame is considered as the whole, as the outer environment. Agent I obtains the optimal control strategy through repeated training, so that the bilevel random decision model can calculate the optimal solution faster and more accurately.

4 SIMULATION

4.1 Basic situation description of test case

In order to verify the effectiveness of the proposed method, this paper chooses the National 863 project "Research and Demonstration of Active Distribution Network Integrating Renewable Energy" as an example to test and analyze the field data of the demonstration project in Hongfeng District, Guizhou Province.

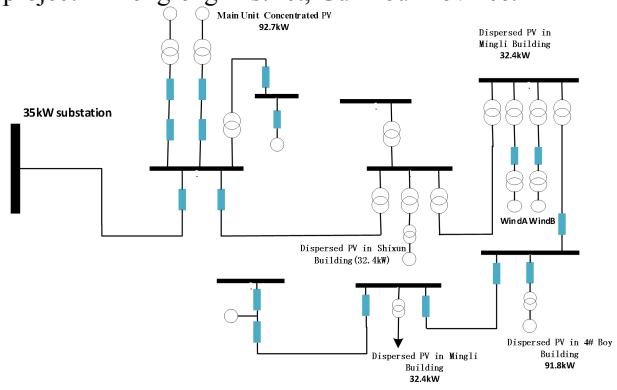

Fig.4 Grid structure of the active distribution network demonstration project

The study mainly focuses on the distributed generators involved in bidding decisions, which are composed of wind power and photovoltaics, as shown in Table 1.

Table 1 Table of the distributed power parameters on shuipei line

Power name	Rated capacity (kW)	Generator Abbreviation
4#1 Ring Main Unit Wind Power A	100	WP1
4#1 Ring Main Unit Wind Power B	100	WP2
2# Ring Main Unit Concentrated PV	92.7	PV1
Dispersed PV in Shixun Building	32.4	PV2
Dispersed PV in 4# Boy Building	91.8	PV3
Dispersed PV in Mingli Building	32.4	PV4

The 24-hour power generation prediction data of the distributed power supply of the hydroponic line derived from the demonstration project active distribution network global energy management system is shown in Fig. 5.

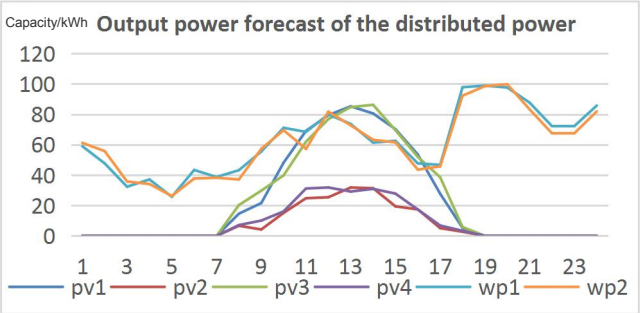

Fig.5 Output power forecast of the distributed power

In this example, the allocation transformations are taken as the load points of participating in the subscription. There are five allocation transformations in the region, which are regarded as the five loads of participating in the subscription. Their names and assumed energy subscription ratios of the day are shown in Table 2.

Table 2 Table of the load information on shuipei line

Name of distribution transformer	Wind power subscription ratio	PV subscription ratio	User abbreviation
Box Transformer Substation in Mingli Building	35%	15%	LD1
#4 Student Building	30%	5%	LD2
Substion in Teaching Building	20%	10%	LD3
1#Box Transformer Substation in Hotel	30%	20%	LD4
Box Transformer Substation in Shixun Substation	25%	30%	LD5

Hypothesis:

The electricity purchase price of the area from the external power grid refers to the grid price of coal-fired units in Guizhou Province, which is 0.3709 yuan/kWh. The penalty for the assessment of the power producer is the same as the purchase price of the external grid, which is 0.3709 yuan/kWh.

The price of wind power benchmark is 0.57 yuan/kWh, the price of PV benchmark is 0.85 yuan/kWh, the subscription fee for electric wind power per user is 0.15 yuan, and the subscription fee for PV is 0.35 yuan. The rest subsidies are borne by the government.

4.2 Result

Using multi-agent bilevel cooperative reinforcement learning, the algorithm can solve the problem. The scalar power allocation in the system can be obtained as shown in Fig. 6. As shown in the label bar, the color blocks of different colors represent the winning power allocation of different generators in this period. OUT refers to the amount of electricity purchased from the external power grid during that period. The second day PV energy structure accounted for 15.14% and wind power 35.00% according to this scheduling scheme. The energy

structure meets the subscription requirements of all users for wind power and LD1, LD2 and LD3 users for PV subscription.

The validity of the model solving algorithm is verified by comparing with the standard particle swarm optimization which solves the benchmark example at the same time. Because there are many optimization variables (144 decision variables), the population size is set to 500 particles. After testing, the learning factor c1c2 is 2, and the inertia coefficient is 0.6-0.3 linear decreasing strategy. The algorithm converges after 500 runs, and the result is the best. The results of optimal power allocation are shown in Fig. 7. We can be seen that the trend of power allocation is consistent with the results of this algorithm.

The expected energy proportion of PSO (Particle swarm optimization) scheduling is 15.22% for PV and 35.00% for wind power, which is very close to the scheduling results of this algorithm. At the same time, considering the objective function minimization problem, the optimal value of particle swarm optimization algorithm is 2062.7988 (with 95% risk confidence), which is slightly larger than the 2062.376 (with 95% risk confidence) obtained by the proposed algorithm, which proves the correctness of this algorithm. At the same time, on the premise of obtaining similar results, the time required for 500 steps of particle swarm optimization iteration is 1175 ms, while the proposed algorithm only needs about 25 ms (even can be reduced to less than 10 ms with relaxed convergence accuracy). This not only proves the correctness of the algorithm in this paper, but also shows that the algorithm can effectively accelerate the speed of model solving.

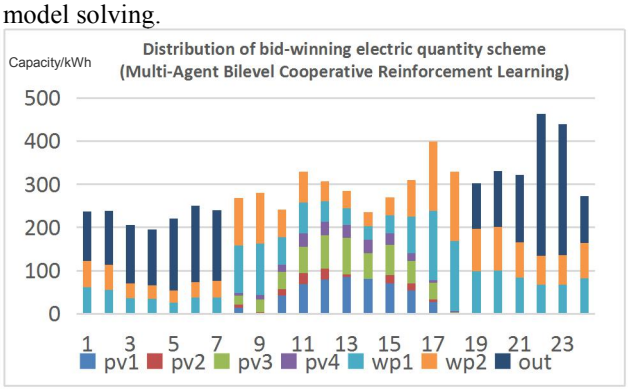

Fig.6 The electricity allocation solved by Reinforcement Learning

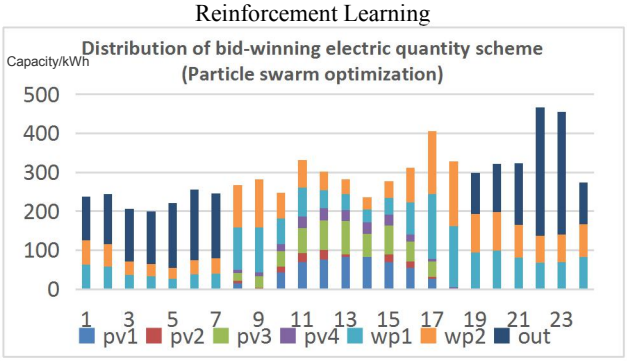

Fig.7 The electricity allocation solved by PSO

5 CONCLUSION

In the context of energy internet, the coordination and complementarity of multiple energy sources is conducive to improving the comprehensive utilization efficiency of energy and the economy of the overall operation of multiple energy systems. Based on the establishment of a two-layer stochastic decision-making optimization model for distributed renewable energy transactions, this paper proposes a multi-agent double-layer cooperative reinforcement learning algorithm. The model and algorithm have the following characteristics:

1) The bilevel decision-making optimization model established in this paper can comprehensively consider the uncertainty scenarios caused by these random variables and make better decisions. Therefore, it is very suitable for the optimization decision-making of distributed generators in this paper.

2) This algorithm is a bilevel reinforcement learning algorithm, which can be well integrated into the two-layer stochastic decision-making optimization model, which provides a new idea for the intensive energy trading decision of information network and energy network in the future.

3) Multi-agent bilevel cooperative reinforcement learning as a multi-agent reinforcement learning algorithm with self-learning and cooperative learning ability, more suitable for solving large-scale distributed access energy with strong randomness and uncertainty. Trading problem. After a certain amount of training update, the algorithm can quickly perform dynamic optimization while ensuring the stability of global convergence.

REFERENCES

[1] YU Shuang, 2014, "A Bidding Model for a Virtual Power Plant Considering Uncertainties", *Automation of Electric Power Systems.* vol. 38, 43-49.

[2] L. Shi, 2014, "Bidding strategy of microgrid with consideration of uncertainty for participating in power market", *International Journal of Electrical Power & Energy Systems*. 59(7), 1-13.

[3] Palma-Behnke R, et al, 2005, "A distribution company energy acquisition market model with integration of distributed generation and load cur-tailment options", *IEEE Transactions on Power Systems.* 20(4), 1718-1727.

[4] Li H, et al, 2007, "A Multiperiod Energy Acquisition Model for a Distribution Company with Distributed Generation and Interruptible Load", *IEEE Transactions on Power Systems*. 22(2), 588-596.

[5] C.Watkins, 1992, "Q-Learning", *Machine Learning.* (11). 279-292

[6] Peng Jing, Williams R J, 1996, "Incremental multi-step Q-learning", *Machine Leaning.* 22. 283-290

[7] Singh S P, 1996, "Reinforcement learning with replacing eligibility trace", *Machine Learning.* 22. 123-158

[8] Richard S.Sutton, Andrew G.Barto, 1998, *Reinforcement Learning: An Introduction,* Cambridge: MIT Press.

[9] Zhang, Q, 2018. "A double deep q-learning model for energy-efficient edge scheduling". *IEEE Transactions on Services Computing*, 1-1.