IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 16, 2021, accepted March 29, 2021, date of publication April 8, 2021, date of current version April 15, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3071803

Research on Cooperation Between Wind Farm and Electric Vehicle Aggregator Based on A3C Algorithm

YANG PAN, WEIYE WANG, YANBIN LI, FENG ZHANG, YANTING SUN, AND DUNNAN LIU
School of Economics and Management, North China Electric Power University (NCEPU), Beijing 102206, China

Corresponding author: Feng Zhang (ncepuzf@ncepu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFE0208100, and in part by the 2018 Key Projects of Philosophy and Social Sciences Research, Ministry of Education, China, under Grant 18JZD032.

ABSTRACT As renewable energy sources such as wind are connected to the grid on a large scale, the safe and stable operation of the power system is facing challenges and the demand for flexibility is becoming increasingly prominent. In recent years, with the advancement of Vehicle-to-Grid (V2G) technology, electric vehicles (EVs) have become a non-negligible flexibility resource for the power system and an emerging path to solve the renewable energy consumption problem. To address the problem of wind farms' difficulty in making profits in the power market, this paper considers the cooperation between wind farms and EV aggregators and uses the levelable characteristics of EVs charging load to ease the anti-peak characteristics of wind power. Given this, this paper proposes a cooperation mode between the wind farm and the Electric Vehicle (EV) aggregator, constructs a cooperation income and income distribution model, and solves the model using the Asynchronous Advantage Actor-Critic (A3C) reinforcement learning algorithm. Finally, based on the simulation analysis of historical data, the following conclusions are drawn: (1) the cooperation between the wind farm and the EV aggregator can effectively mitigate the negative impact of the anti-peak characteristics of wind power on profitability and achieve an increase in overall economic benefits; (2) the income distribution based on the Shapley value method ensures that the respective income of the wind farm and the EV aggregator increase after cooperation, which is conducive to the promotion of the willingness of both parties to cooperate; (3) the A3C reinforcement learning algorithm is applied to solve the model with good convergence to achieve fast and continuous intelligent pricing decisions for EV aggregators, thus optimizing the charging schedule of EVs promptly.

INDEX TERMS Electric vehicle aggregator, wind farm, A3C algorithm, charging service pricing.

I. INTRODUCTION

With the increasing depletion of fossil energy and the increasing seriousness of climate and environmental problems, the traditional energy production and consumption methods are unsustainable, and the large-scale development and utilization of renewable energy have become an important strategy for sustainable energy development in many countries. Australia plans to achieve 100% renewable energy supply by 2050; the European Union will reduce greenhouse gas emissions by 2030 from the original 40% to 55%; the United States is committed to achieving carbon-free power generation by 2035 and "100% clean energy consumption" by 2050 [1]–[3]. China is committed to achieving carbon neutral by 2060, and a high percentage of renewable energy connected to the grid will be an important feature of China's future power system [4].

However, due to the uncontrollable power output, the current renewable energy consumption in China faces a severe situation, such as the intermittent and anti-peak characteristics of wind power, which brings challenges to the safe and stable operation of the power system and puts forward higher requirements for power system flexibility [5], [6]. In recent years, as electric vehicles (EVs) are connected to the grid on a large scale, the energy storage characteristics of their on-board batteries have gradually been emphasized as an

The associate editor coordinating the review of this manuscript and approving it for publication was Kok-Lim Alvin Yau.

effective path to enhance the flexibility of the power system [7]–[9].

At present, using the flexibility of EVs' charging load to solve the problem of wind power consumption difficulties has become a hot topic of concern for many scholars [10]. Thereinto, in the power market, scheduling charging plans for EVs through EV aggregators can solve the problem of bidding errors by wind farms due to uncertainty in wind power output. Divya *et al.* [11] proposed a rational market mechanism that uses the energy storage characteristics of electric vehicles to reduce the additional bidding costs due to wind power forecast errors. Vaya and Andersson [12] examined the problem of self-dispatch for EV aggregators purchasing power in the day-ahead market and providing balancing services for forecast errors of wind farms. Hu *et al.* [13] proposed optimal operating strategies for EV aggregators in the spot and ancillary service markets in the power system containing a high percentage of wind power. Also, EVs and wind turbines can form virtual power plants (VPPs) to participate in power market transactions together. Alahyari *et al.* [14] discussed trading strategies for VPP operators with wind turbines and parking lots for EVs to participate in both the power and reserve markets. Vasirani *et al.* [15] developed a profit model for a VPP containing wind turbines and EVs, solved it using linear programming, and the data simulation showed that the model has better economic efficiency for participation in the power market. Massive research has also been conducted on the co-dispatching method of EVs and wind turbines. Kou *et al.* [16] proposed a hierarchical stochastic control model for EVs and wind power synergy in the microgrid to achieve supply and demand power balance. Zhu *et al.* [17] developed a joint scheduling model and solved it using an improved multi-objective decomposition-based evolutionary (IMOEA/D) algorithm. Zhao *et al.* [18] proposed an economic dispatch model for wind turbines and EVs and developed the IPPSO algorithm based on improved particle swarm optimization and interior point method for solving. Korkas *et al.* [19] proposed an intelligent optimization method based on multi-modal approximate dynamic programming (MM-ADP) to achieve optimal charging and discharging vehicle scheduling for grid-connected charging stations.

Based on economics, the realization of the collaborative dispatch of EVs and wind turbines is inseparable from the pricing decision of charging services for EV customers [20]. Li and Ouyang [21] studied the factors influencing the pricing of charging services, such as financial subsidies, operating costs, and charging revenues. Zhuang *et al.* [22] analyzed the costs and profits of EV aggregators and developed a cost-effective pricing methodology for charging services. Wu *et al.* [23], Tushar *et al.* [24], Yang *et al.* [25] studied the dynamic pricing problem of EV aggregators from a game-theoretic perspective, achieving goals such as maximizing profits or providing ancillary services to the grid. Zhang *et al.* [26] developed a comprehensive model from the perspective of customer satisfaction to derive a time-sharing pricing strategy for different types of customers. Chen *et al.* [27] proposed a dynamic pricing model based on the clustering algorithm that considers unknown information such as grid voltage distribution dynamics and actual random arrival and departure times of EVs. Nie *et al.* [28] proposed a new smart city modeling approach based on EV travel charging networks, applying a multi-area adaptive pricing scheme to effectively solve the problems posed by complex transportation networks and large-scale EV integration.

Based on the aforementioned research, extensive previous research has been conducted on the issue of wind power consumption by EVs and charging pricing for EV customers, which has laid the theoretical foundation for the research in this paper, but several problems remain unresolved: (1) Most of the above studies started from the perspective of electric vehicles to promote wind power consumption and did not consider how to solve the problem of wind power being difficult to make a profit in the power market due to its anti-peak characteristics. (2) The above studies have dealt with the strategies of wind power and EV participation in the power market as well as collaborative dispatching models and methods, but less research has been conducted on the cooperation mode of the wind farm and the EV aggregator and the distribution of cooperation income. (3) The above studies have used influence factor analysis, cost-benefit analysis, game theory, and the clustering algorithm to study the pricing problem of EV charging services, but in the face of the complexity and variability of the external environment, EV aggregators need to make continuous and timely dynamic pricing decisions to maximize benefits.

Therefore, this paper focuses on the problem that wind power is not profitable in the power market due to its anti-peak characteristics, and considers the cooperation between the wind farm and the EV aggregator and the pricing of EV charging services, and researches how the two parties cooperate and allocate their income to reduce the negative impact of wind power's anti-peak characteristics and achieve economic benefits and the willingness of both parties to cooperate; how the EV aggregator make dynamic pricing decisions for charging services, thereby optimize the charging schedule of EVs promptly and maximize the income of cooperation. Asynchronous Advantage Actor-Critic (A3C) is a kind of reinforcement learning algorithm based on probability selection, which includes data perception ability. It has been widely used in the field of decision making due to its characteristics of fast decision-making speed, strong robustness, asynchronous parallel processing and so on [29]–[31]. Therefore, in a future where a high percentage of renewable energy and EVs are connected to the power system, this paper can provide a reference for cooperation between EVs and wind power, and even renewable energy, as well as for pricing decisions for EV aggregators.

The possible innovations of this paper are listed as follows:

(1) A cooperation mode between the wind farm and the EV aggregator is proposed, in which the wind farm is given

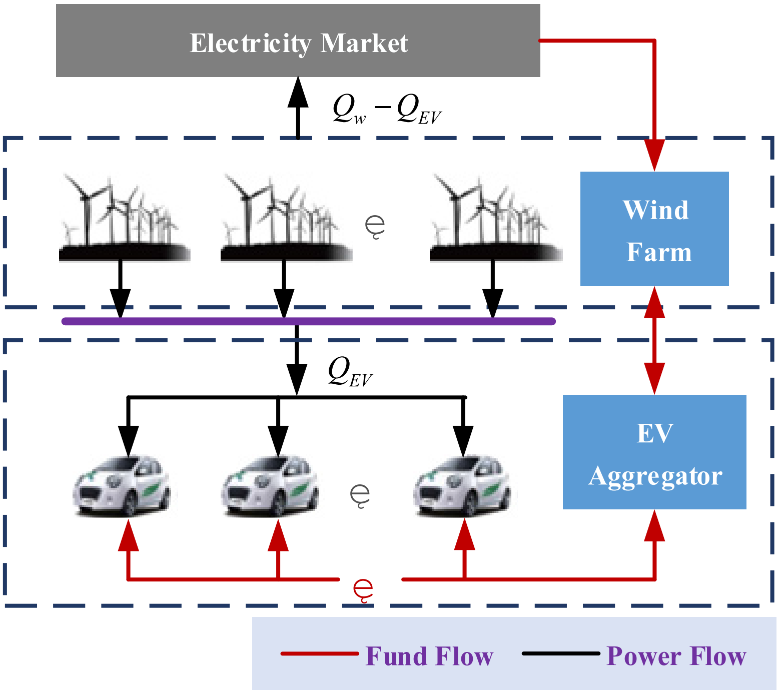

FIGURE 1. Cooperation mode between the wind farm and the EV aggregator.

priority to supply power to EVs, and the remaining power is traded in the power market; the EV aggregator guides customers to charge in an orderly manner through pricing to ease the negative impact of the anti-peak characteristics of wind power on profitability.

(2) An income allocation model is constructed based on the Shapley value method. Under the goal of maximizing cooperative income, the income is redistributed to the wind farm and the EV aggregator at the micro level to promote the willingness of both parties to cooperate.

(3) The A3C reinforcement learning algorithm is applied to the pricing decision of the EV aggregator to achieve fast and continuous pricing decisions for EV charging services, reflecting the foresight of AI application to solve pricing problems.

The rest of this paper is organized as follows. In Section 2, the cooperation mode between the wind farm and the EV aggregator is proposed, and the income and income distribution model are constructed. In Section 3, the A3C reinforcement learning algorithm is applied to EV charging service pricing. In Section 4, a case study is conducted to demonstrate the effectiveness of the proposed cooperation model and the A3C reinforcement learning algorithm. Section V draws the main conclusions and indicates the next research direction.

II. MODEL FORMULATION

A. PROBLEM DESCRIPTION

Wind power has anti-peak characteristics, i.e. wind power often peaks at night, when the system load is low, the power is oversupplied and the market electricity price is low [15]. Therefore, wind farms are at a competitive disadvantage in the power market, and with the intermittent and uncontrollable nature of wind power, it is more difficult for wind farms to achieve profitability. To promote wind power consumption, many countries implement higher feed-in tariffs than market tariffs for wind farms. But in the long run, the higher cost of power generation goes against the principle of the economics of power system operation and hinders the effective allocation of power resources by the market.

To solve the above problem, we design a cooperation mode between the wind farm and the EV aggregator by using the levelable charging load characteristics of EVs to ease the anti-peak characteristics of wind power. As shown in Figure 1, the wind farm gives priority to supply power to the EV aggregator, and sells leftover to the power market; the EV aggregator provides charging services for the EVs, and charges power purchase costs plus service fees; finally, the wind farm and the EV aggregator share the cooperation income.

B. INCOME CALCULATION MODEL

1) NON-COOPERATIVE INCOME CALCULATION

The study in this paper focuses on the cooperative income of the wind farm and the EV aggregator, so the generation cost of the wind farm is not considered. The wind farm sells all of its power in the power market when it does not cooperate with the EV aggregator, so the non-cooperative wind farm's income can be calculated as follows:

$$R_{wind} = \sum_{t=1}^{T} P_m^t Q_w^t \tag{1}$$

where R_{wind} represents the income of the non-cooperative wind farm; P_m^t is market electricity price at time t; Q_w^t is electricity sold by the wind farm in the power market at time t.

Likewise, this paper considers only the electricity purchase cost of the EV aggregator. When the EV aggregator is non-cooperative, it purchases electricity from the grid and then sells to EV customers at the retail price of power purchase costs plus service fees. Thus, the income of the non-cooperative EV aggregator can be calculated as follows:

$$R_{EV} = \sum_{t=1}^{T} \left(\left(P_g^t + P_s^t \right) Q_{EV}^t - P_g^t Q_{EV}^t \right) = \sum_{t=1}^{T} P_s^t Q_{EV}^t \quad (2)$$

where R_{EV} represents the income of the non-cooperative EV aggregator; P_g^t is the price for EV aggregator purchasing electricity from the grid at time t; P_s^t is the additional service rate charged by the EV aggregator at time t; Q_{EV}^t is the charging load of EVs at time t.

2) COOPERATIVE INCOME CALCULATION

When the wind farm and the EV aggregator cooperate, the wind farm gives priority to supply power to EVs and sells the remaining power in the power market; the EV aggregator guides EV customers to orderly charging by pricing the service fee. Therefore, the income of the cooperation between the wind farm and the EV aggregator can be calculated as follows:

$$R_{co} = \sum_{t=1}^{T} \Big(\left(P_g^t + P_s^t \right) \left(Q_{EV}^t + \Delta Q^t \right) + P_m^t \left(Q_w^t - Q_{EV}^t - \Delta Q^t \right) \Big) \quad (3)$$

where R_{co} represents the income of the cooperation between the wind farm and the EV aggregator; ΔQ^t is the amount of change in charging load for the demand response of the EV customers to the charging service pricing at time t.

EV customers adjust their charging demand based on the change in the sum of EV aggregator service fees and power purchase costs. EV charging demand has a large degree of uncertainty and regularity; the uncertainty comes from the randomness of charging time and charging power of individual EV customer, and the regularity comes from the overall trend of charging demand of large-scale EVs. In this paper, the charging uncertainty of individual EV customer is not considered, and based on the price elasticity of demand theory of microeconomics [20], EV customers' charging demand in period T is adjusted as in follows:

$$E = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1n} \\ e_{21} & e_{22} & \cdots & e_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ e_{n1} & e_{n2} & \cdots & e_{nn} \end{bmatrix} \quad (4)$$

$$T = nt \quad (5)$$

The elasticity coefficients are described as follows:

$$e_{ij} = \begin{cases} \dfrac{\frac{\Delta Q_i}{Q_i}}{\frac{\Delta P_i}{P_i}}, & \&i = j \\[2ex] \dfrac{\frac{\Delta Q_i}{Q_i}}{\frac{\Delta P_i}{P_i}}, & \&i \neq j \end{cases} \quad (6)$$

In this paper, we introduce value function [32] to study the relationship between charging pricing and load, and use customers' load data and charging price changes to estimate unknown parameters to calculate the elasticity coefficients e_{ij}. Based on the price elasticity matrix shown in equation (4), the amount of change in charging demand by EV customers, i.e., the amount of change in electricity sold by the EV aggregator ΔQ in period T, can be calculated as follows:

$$\Delta Q = \begin{bmatrix} \Delta Q_1 \\ \Delta Q_2 \\ \cdots \\ \Delta Q_t \end{bmatrix} = \begin{bmatrix} Q_1 & 0 & \cdots & 0 \\ 0 & Q_2 & \cdots & 0 \\ \cdots & 0 & \cdots & \cdots \\ 0 & 0 & \cdots & Q_n \end{bmatrix} * \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1n} \\ e_{21} & e_{22} & \cdots & e_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ e_{n1} & e_{2n} & \cdots & e_{nn} \end{bmatrix} * \begin{bmatrix} \frac{\Delta P_1}{P_1} \\ \frac{\Delta P_2}{P_2} \\ \cdots \\ \frac{\Delta P_n}{P_n} \end{bmatrix} \quad (7)$$

C. INCOME OPTIMIZATION MODEL

This paper aims to maximize the income of cooperation between the wind farm and the EV aggregator, so the income optimization model is described as follows:

$$max \ \ \Delta R = R_{co} - \left(R_{wind} + R_{EV} \right) \quad (8)$$

$$s.t. \ \sum_{t=1}^{T} Q_{EV}^t = \sum_{t=1}^{T} \left(Q_{EV}^t + \Delta Q^t \right) \quad (9)$$

$$\sum_{t=1}^{T} P_s^t \left(Q_{EV}^t + \Delta Q^t \right) \leq \sum_{t=1}^{T} P_s^t Q_{EV}^t \quad (10)$$

$$P_s^{min} \leq P_s^t \leq P_s^{max} \quad (11)$$

$$Q_{EV}^{min} \leq Q_{EV}^t + \Delta Q^t \leq Q_{EV}^{max} \quad (12)$$

The objective function (8) is the maximization of the incremental income of the wind farm and the EV aggregator when they cooperate compared to their respective non-cooperation. Constraint (9) is that the total charging load of EVs remains unchanged during period T, to ensure the daily driving demand of EV customers. Constraint (10) is to ensure that the cost of EV customers does not increase, to prevent merchants from pricing arbitrarily to increase their profits. Constraint (11) is to set the upper and lower price limits for the service fee at time t. Constraint (12) is to set upper and lower limits for the charging load of EVs at time t.

In this paper, we choose the maximization of incremental income of cooperation between the wind farm and the EV aggregator as the objective function instead of the maximization of cooperative income, which raises the efficiency of solving the model and also argues the necessity of cooperation between them more adequately. Meanwhile, setting the constraints that the total charging load remains unchanged and the charging cost of customers does not increase is conducive to arguing that the increase of cooperation income comes from the improvement of economic efficiency, rather than harming customers' charging demand and interests. Also, setting the upper and lower limits for service fees and charging load is based on the actual situation of price regulation and the limited number of the EV aggregator's charging piles.

The incremental income optimization problem after the cooperation between the wind farm and the EV aggregator is essentially the EV aggregator's pricing decision problem for charging services to EV customers i.e., the EV aggregator guides the orderly charging of EVs i.e., adjusting ΔQ, by deciding on the service fee pricing P_s to EV customers, to optimize the incremental income of cooperation between the two. Since P_s is not a continuous variable, and there is no linear relationship between the incremental income ΔR and the decision variable P_s, the incremental income optimization problem of cooperation between the wind farm and the EV aggregator studied in this paper is nonconvex and nonlinear.

D. INCOME DISTRIBUTION MODEL

In the cooperation between the wind farm and the EV aggregator, how the cooperation income is distributed is directly related to the conclusion of their cooperation. In this paper, the Shapley value method is applied to study the income distribution problem of cooperation between the wind farm and the EV aggregator. The Shapley value method is an equitable distribution method for m-player cooperation. Its core idea is to distribute the income of the participants according to their contributions to the alliance, and the more the contributions, the more the income [33].

Let a subset $S \subseteq M$ of any non-empty set $M = d\{1, 2, \ldots, m\}$ of participants, called alliance. Use the Shapley value method to calculate the income φ assigned to the participant i, as follows:

$$\varphi_i = \sum_{S} \omega\left(|S|\right)\left(\upsilon\left(S\right) - \upsilon\left(S - \{i\}\right)\right) \tag{13}$$

$$\omega\left(|S|\right) = \frac{(m - |S|)!\,(|S| - 1)!}{m!} \tag{14}$$

where $|S|$ is the number of participants in subset S; $\upsilon\left(S\right)$ is the income of the alliance that includes participant i; $\upsilon\left(S - \{i\}\right)$ is the income of the alliance that does not include participant i; $\omega\left(|S|\right)$ is the weighting factor; $m!$ is the number of possible permutations of all participants in the cooperation.

Denote the EV aggregator and the wind farm as 1 and 2, respectively, and $m = 2$. $\upsilon\left(1\right)$ and $\upsilon\left(2\right)$ represents the respective non-cooperative income of the EV aggregator and the wind farm, $\upsilon\left(\{1, 2\}\right)$ denotes the total cooperative income of them. Equations (13) and (14) calculate the respective income allocated to the wind farm and the EV aggregator when they cooperate can be obtained.

III. A3C REINFORCEMENT LEARNING ALGORITHM

Since the objective function of this paper to maximize the incremental income of cooperation between the wind farm and the EV aggregator is nonconvex and nonlinear, the algorithmic complexity of solving the global optimum is exponential (NP-hard). Despite the difficulty in solving the global optimum, nonconvex optimization problems can generally be solved by intelligent optimization algorithms such as genetic algorithm and pattern search for relatively optimal solutions. However, traditional intelligent optimization algorithms are only suitable for solving static planning problems, i.e., the environment is static in the decision process. In the research scenario of this paper, the environment in which the EV aggregator is located is dynamic, i.e., the wind power output and market electricity price are dynamically and continuously updated, which cannot be accurately modeled. Therefore, it is difficult to realize the continuous dynamic pricing decision of the EV aggregator with traditional intelligent optimization algorithms. This is where reinforcement learning, which can make dynamic decisions without modeling the environment, comes into play. Reinforcement learning only requires an agent to find the relatively optimal policy based on the reward value changes through continuous interaction with the environment after defining the reward function. Therefore, in this paper, the incremental income optimization problem of the EV aggregator cooperation with the wind farm through successive dynamic pricing decisions is suitable to be solved by reinforcement learning methods.

Compared with general reinforcement learning, Asynchronous deep reinforcement learning is an integrated algorithm that combines the perceptual capabilities of deep learning for high-dimensional data with the decision-making capabilities of reinforcement learning [29], [34], [30]. Among the asynchronous deep reinforcement learning methods, A3C performs best for task control in all types of action spaces, merging two types of reinforcement learning algorithms based on values (Q-learning) and action probabilities (Policy Gradients). A3C's optimization model based on reward value and its ability to rapidly process high-dimensional data can quickly guide the EV aggregator in pricing service fees based on wind power output forecast information and market electricity price information, optimize the charging schedule for EVs promptly, and increase the income from the cooperation between the EV aggregator and the wind farm.

A. ALGORITHM PRINCIPLE

Reinforcement learning algorithms are methods by which an agent learns through "trial and error", i.e., it interacts with its environment to obtain reward and punishment information and thereby adjusts its behavior. A3C is a reinforcement learning algorithm based on Actor-Critic, whose agent consists of two parts: Actor is responsible for generating actions and interacting with the environment; Critic is responsible for evaluating the Actor's performance and guiding the Actor's actions in the next stage.

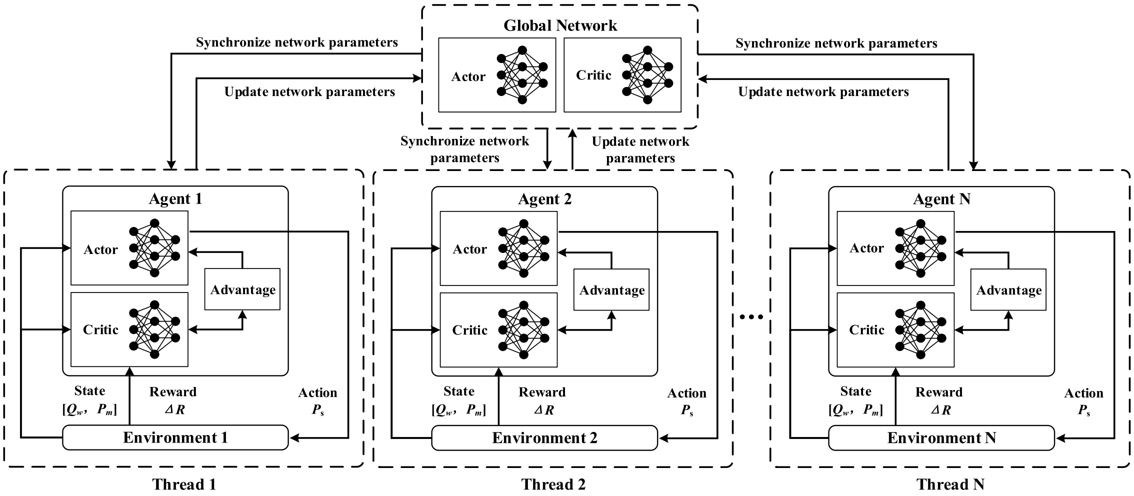

FIGURE 2. Asynchronous training framework for A3C algorithm.

Compared with Actor-Critic, the optimization of A3C is reflected in the 3 aspects: optimization of network structure, asynchronous training framework, and optimization of Critic evaluation points. First, A3C optimizes the network structure by putting the two networks Actor and Critic together, i.e., unified input state S, Critic output state value V, and Actor output corresponding policy π.

Second, A3C builds an asynchronous training framework, which consists of a global network and n worker threads, both of which contain Actor and Critic parts. Each thread interacts with the environment individually to obtain the empirical data, and these threads run independently without interfering with each other. After each thread has interacted with the environment for a certain amount of data, it calculates the gradients of the loss function of the neural network in its thread, but these gradients do not update the neural network in its thread but go to update the global network, i.e., the n threads respectively update the parameters of the global network using the accumulated gradients. The public part of the network model is to be optimized, while the network models in the threads are mainly used to interact with the environment.

Finally, A3C uses N-step sampling to accelerate convergence with an advantage function expressed as:

$$A(S,t) = R_t + \gamma R_{t+1} + \ldots \gamma^{n-1} R_{t+n-1} + \gamma^n V(S') - V(S) \quad (15)$$

For the loss function of Actor and Critic, A3C and Actor-Critic are the same, with one optimization being the addition of the entropy term of the policy π with coefficient c in the loss function of the Actor-Critic policy function, i.e., the gradient update of the policy parameters compared with Actor-Critic is optimized as:

$$\theta = \theta + \alpha \nabla_\theta log\pi_\theta(s_t, a_t) A(S,t) + c\nabla_\theta H(\pi(S_t,\theta)) \quad (16)$$

In this paper, we study how the EV aggregator cooperating with the wind farm can conduct service fee pricing to guide customers to charge in an orderly manner, thereby increasing the income from the cooperation. Therefore, the agent in reinforcement learning is mainly the EV aggregator, which contains both Actor and Critic parts. Based on the scenario of cooperation between the wind farm and the EV aggregator in the power market, the input states S of Actor and Critic include wind power output Q_w and market electricity price P_m, the corresponding policy π output by Actor is the service fee pricing by EV aggregator P_s, and the state value V output by Critic is the incremental income ΔR in period T for the cooperation between the wind farm and the EV aggregator.

B. ALGORITHM FLOW

The asynchronous training framework of the A3C algorithm for the research scenario in this paper is shown in Figure 2, where the Agent within each thread interacts with the Environment, i.e., the Actor selects a policy (service fee P_s) based on a probability distribution, the Critic evaluates the Actor's policy, then the Actor adjusts the probability distribution of the policy based on the Critic's evaluation value; the Environment inputs the policy adopted by the agent, and outputs the reward obtained by the agent (the incremental income of cooperation ΔR in period T) and the state in period $T + 1$ (wind power output Q_w and market electricity price P_m). The agent of each thread is trained independently, the gradients of network parameters are calculated in parallel, and the network parameters are updated; finally, the network parameters of each thread are updated to the global network asynchronously to eliminate the correlation between training data and achieve better and faster convergence of the algorithm.

Since A3C is asynchronous and multi-threaded, the specific algorithm flow for each thread is shown in Table 1.

IV. SIMULATION

A. DATA DESCRIPTION

In this paper, 150 days of real data from a typical region in China are selected for simulation, including daily wind power output data, charging load data of EVs and market

TABLE 1. Algorithm flow for each thread.

```
Algorithm flow:
1. Input: global network's parameters $\theta, \omega$; this thread's parameters $\theta', \omega'$; global shared iteration rounds $T$; global maximum of iterations rounds $T_{max}$; the maximum length of the single iteration time series in the thread $t_{local}$; the state feature dimension $n$; the action set $A$; the step size $\alpha, \beta$; the entropy coefficient $c$; the decay factor $\gamma$
2. Update time series $t=1$
3. Reset the amount of gradient updating for Actor and Critic: $d\theta \leftarrow 0$, $d\omega \leftarrow 0$
4. Synchronize parameters from the global network to the neural network in this thread: $\theta' = \theta, \omega' = \omega$
5. $t_{start}=t$, initialized state $s_t$
6. Select out action $a_t$ based on the policy $\pi(a_t|s_t;\theta)$
7. Execute action $a_t$ to get reward $r_t$ and new state $s_{t+1}$
8. $t \leftarrow t+1$, $T \leftarrow T+1$
9. If $s_t$ is terminated, or $t-t_{start}==t_{local}$, then go to step 10, otherwise, go back to step 6
10. Compute $Q(s,t)$ for the last time-series position $s_t$:
        $Q(s,t)=\begin{cases} 0 & \text{terminal state} \\ V(s_t,\omega') & \text{none terminal state, bootstrapping} \end{cases}$
11. for $i \in (t-1, t-2, \ldots t_{start})$:
    1) Calculate $Q(s,i)$ for each moment: $Q(s,i)=r_i+\gamma Q(s,i+1)$
    2) Accumulate local gradient updates for Actor:
        $d\theta \leftarrow d\theta + \nabla_{\theta'} \log\pi_{\theta'}(s_i,a_i)\left(Q(s,i)-V(s_i,\omega')\right)+c\nabla_{\theta'} H(\pi(s_i,\theta'))$
    3) Accumulate local gradient updates for Critic:
        $d\omega = d\omega + \frac{\partial\left(Q(s,i)-V(s_i,\omega')\right)^2}{\partial\omega'}$
12. Update parameters of the global network:
        $\theta=\theta-\alpha d\theta$, $\omega=\omega-\beta d\omega$
13. If $T>T_{max}$, the algorithm flow ends and outputs global network's parameters $\theta, \omega$, otherwise, go back to step 4r
```

electricity price data, etc., to verify the feasibility of the cooperation mode proposed and the validity of the model. Meanwhile, the EV aggregator in the region purchases electricity from the grid on a peak-to-valley tariff, with the peak hour (13:00-22:00) tariff being RMB 1289.3/MWh, the flat hour (8:00-13:00 and 22:00-1:00) tariff being RMB 873.1 /MWh, and the valley hours (1:00-8:00) tariff being RMB 457/MWh; the service fee charged by the EV aggregator is a fixed price of RMB 500/MWh. Besides, due to the limited number of charging piles of EV aggregator, the charging power range for accessing EVs per moment is 0∼7000MWh; for price regulation, the pricing range of service fee for EV aggregator is set at 0∼1000 RMB/MWh.

The optimization process of the A3C reinforcement learning algorithm is realized through the joint optimization of Actor and Critic neural networks. It takes about 30 seconds per episode to train the Actor and Critic neural networks, but the agent can output the decision in less than 1 second according to the environment after the training is completed. We use the A3C algorithm to solve the model, simulate the above real data, train 2000 episodes, which takes about 17 h. The training of each episode consists of 150d data, the batch size is 32, the number of hidden nodes of the neural network is 500. The simulation environment is Intel core i7-8700@3.20GHz, 6 cores and 12 threads, memory 16GB, software configuration Python 3.7.0, TensorFlow 2.2.0. Table 2 shows the A3C algorithm parameters setting.

TABLE 2. The A3C algorithm parameters setting.

parameters	symbols	numerical
global shared iteration rounds	T	2000
global maximum of iterations rounds	T_{max}	150
the maximum length of the single iteration time series in the thread	t_{local}	150
the state feature dimension	n	0.01
the step size	α	0.002
the step size	β	0.007
the entropy coefficient	c	0.005
the decay factor	γ	0.99

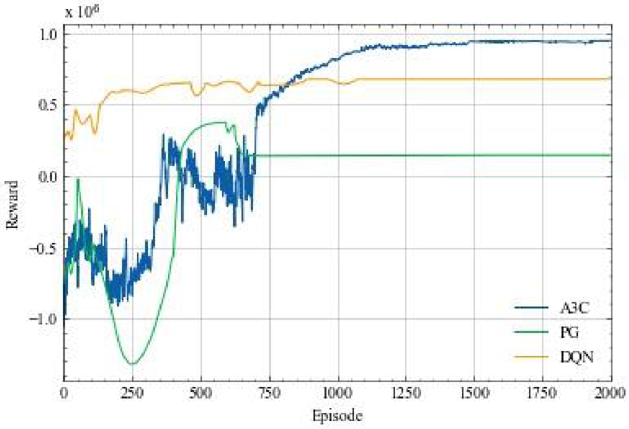

FIGURE 3. Search process.

B. RESULTS ANALYSIS

1) EVALUATION OF ALGORITHM CONVERGENCE

To compare and analyze the convergence of the A3C algorithm, we performed simulations using the Policy Gradient (PG) algorithm and the Deep Q-Learning (DQN) algorithm based on the same data. Figure 3 shows the convergence of the three types of reinforcement learning algorithms, i.e., the learning capability of the agent and its obtained reward value, which in this paper refers to the pricing decision capability of the EV aggregator and the income of cooperation with the wind farm. It can be seen that in the same training time, the DQN algorithm has the highest initial reward value in the first episode, but the subsequent growth space is small and the training is inefficient; the subsequent growth of the PG algorithm is significant but fluctuates greatly and does not converge at the highest point, and the training effect is unstable.; and the reward value of the A3C algorithm that finally converges is the highest. When simulating with the A3C algorithm, the reward value of the agent also fluctuates a lot and even appears negative, but this is because of the large random factor added at the beginning of training to expand the action range of the policy network. after 750 episodes,

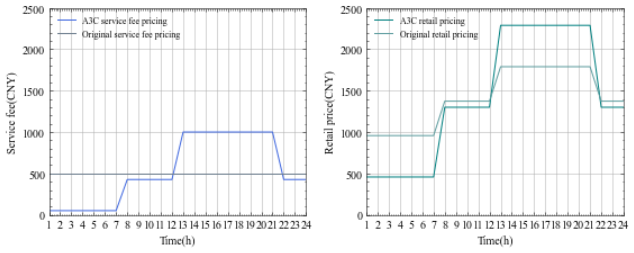

FIGURE 4. Pricing to EV customers.

the reward value of the agent tends to grow steadily, and the growth rate slows down. Finally, it converges to a stable level after 1500 episodes. It is proved that the A3C reinforcement learning algorithm has a good convergence ability to solve the model.

2) EFFECT ANALYSIS OF COOPERATION MODE

This paper analyzes the effect of the proposed cooperation mode between the wind farm and the EV aggregator in terms of three changes in pricing strategies of EV aggregator, charging behavior of EV customers, and the amount of electricity sold by the wind farm in the power market.

The original pricing of the aggregator to EV customers is to add a fixed service fee to the cost of electricity purchase, while the application of the A3C algorithm is to dynamically adjust the pricing of service fees according to the price of electricity sold by the wind farm in the power market, to guide EV customers to charge in an orderly manner and increase cooperation income. The dynamic pricing of a typical day is selected to compare with the original pricing as shown in Figure 4. After applying the A3C algorithm to dynamically adjust the service fee pricing, the original fixed service fee becomes time-sharing pricing, thus exacerbating the peak-to-valley difference in the retail pricing of EV aggregator charging services.

According to the price elasticity of demand theory of microeconomics, after the EV aggregator changes the pricing of charging service, customers will adjust their EV charging demand accordingly. A typical daily charging load of customers is selected as shown in Figure 5. After the EV aggregator applies the A3C algorithm to adjust the service fee pricing, customers tend to increase the charging load from 1:00 to 7:00 and decrease the charging load from 12:00 to 21:00, which is in line with the economic principle that customers charge more when the price is lower in the valley hours and less when the price is higher in the peak hours, proving the validity of applying the A3C algorithm pricing.

The levelable charging load characteristics of EVs are used to ease the anti-peak characteristics of wind power, which is the basis for cooperation between the wind farm and the EV aggregator. The electricity sales and market price of the wind farm in the power market on a typical day are shown in Figure 6. After dynamic pricing by EV aggregator, from 1:00 to 8:00, the power market price is low, and wind power sales in the power market tend to decrease; from 12:00 to 22:00, the power market price is high, and wind power sales in the power market tend to increase. The negative impact of wind power's anti-peak characteristics on income is mitigated, which proves the effectiveness of the cooperation mode proposed in this paper.

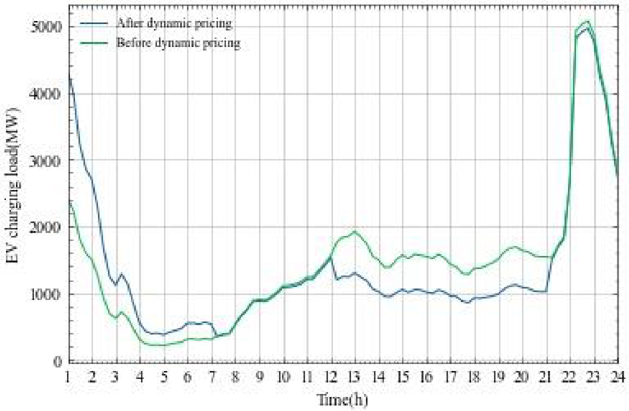

FIGURE 5. Charging load of EV customers.

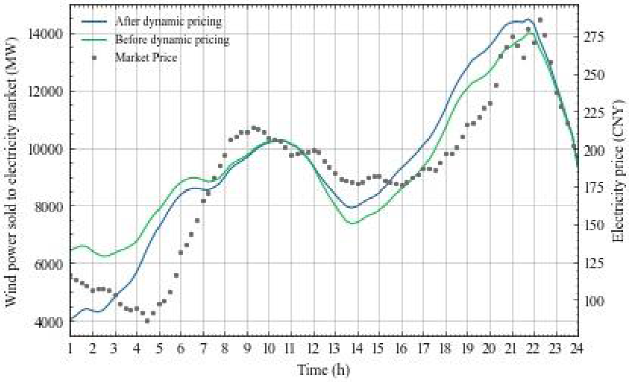

FIGURE 6. Electricity sold by the wind farm to power market.

3) RESULT ANALYSIS OF INCOME DISTRIBUTION

The increase of income is a prerequisite for the willingness to cooperate between the wind farm and the EV aggregator. A comparison of the total income of cooperation and non-cooperation is shown in Figure 7, which shows that the total income of cooperation is higher than the total income of non-cooperation every day.

The Shapley value method is used to allocate the total cooperative income between the wind farm and the EV aggregator, and the income shared by the wind farm and the EV aggregator is compared with their non-cooperative income, respectively. As shown in Figure 8, after their cooperation, the income shared by the wind farm is higher compared to its non-cooperative income, and the income shared by the EV aggregator is higher compared to its non-cooperative income.

Finally, Table 3 gives specific values for the respective and total income of the wind farm and the EV aggregator for 150 days in the non-cooperative and cooperative scenarios. Overall, the respective income and total income increase

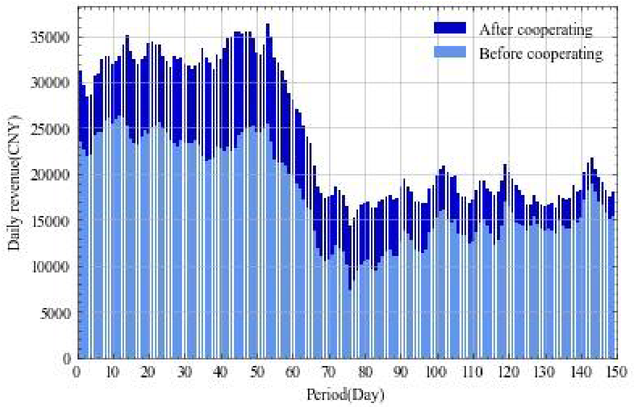

FIGURE 7. Total income comparison.

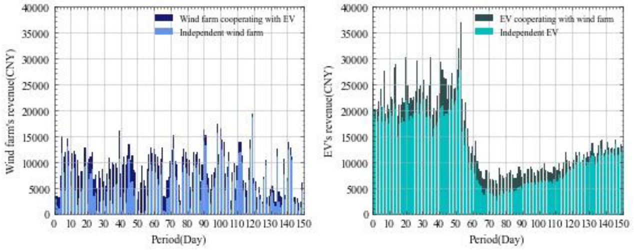

FIGURE 8. Respective income comparison.

TABLE 3. Income comparison.

Scenarios	The income of wind farm /CNY	The income of EV aggregator /CNY	Total income /CNY
Non-cooperation	849938.01	1803291.91	2640314.12
Cooperation	1319624.94	2280925.38	3595581.06

significantly after their cooperation compared to their operating independently. It is calculated that the income of the wind farm increases by 55.26%, the income of the EV aggregator increases by 26.49%, and the total income increases by 36.18% after the cooperation.

In summary, the increase in the total income from the cooperation argues for the necessity of cooperation between the wind farm and the EV aggregator from the perspective of maximizing economic benefits. The increase in the respective income shows the feasibility of cooperation between the wind farm and the EV aggregator, which is profitable can lead to the willingness of both parties to cooperate.

V. CONCLUSION

In this paper, a cooperation mode between the wind farm and the EV aggregator is proposed to address the problem that wind power is not profitable in the power market due to its anti-peak characteristics. Next, the cooperative income optimization and income distribution model is built. Then, the A3C reinforcement learning algorithm is applied to solve the model, i.e., to price EV charging services. Finally, a typical region in China is selected as a case sample. The main conclusions drawn are as follows:

(1) The cooperation mode between the wind farm and the EV aggregator proposed in this paper is feasible. The cooperation mode makes use of the levelable characteristics of EV charging load to effectively ease the anti-peak characteristics of wind power and realize the improvement of economic benefits, which is in line with the development trend of the power system under energy transition and provides a reference for the cooperation between EV and renewable energy.

(2) The income distribution based on the Shapley value method ensures reasonable income for the wind farm and the EV aggregator, i.e., based on the increase in cooperation income, the increase in their respective income is also achieved, which is conducive to the willingness of both parties to cooperate from the perspective of maximizing individual income.

(3) The A3C reinforcement learning algorithm solves the model with good convergence, stability, and timeliness. The algorithm extracts high-dimensional data information features of wind power output and power market, and hands over to the agent to execute multi-action optimization decision to realize intelligent pricing decision of EV aggregator charging service and optimize charging schedule of EVs in time.

The research in this paper focuses on the cooperation mode between the wind farm and the EV aggregator, income distribution and pricing decision behavior of the EV aggregator, and the discharging behavior of EVs is not considered in the model for the time being. In the next research, V2G technology of EVs will be introduced to further explore the cooperation between EVs and renewable energy sources such as wind power and PV as well as the coordination of internal interests.

REFERENCES

[1] *Biden's Clean Energy Revolution and Environmental Justice Plan*. Accessed: Nov. 9, 2020. [Online]. Available: https://news.solarbe.com/202011/09/332049.html

[2] *Australia Aims to Achieve 100% Renewable Energy by 2050*. Accessed: Apr. 23, 2015. [Online]. Available: http://guangfu.bjx.com.cn/news/20150423/611142.shtml

[3] *The European Union Plans to Increase ts Greenhouse Gas Reduction Targets*. Accessed: Sep. 29, 2020. [Online]. Available: http://paper.people.com.cn/rmrb/html/2020-09/29/nw.D110000renmrb_20200929_3-17.htm

[4] *China Aims to Achieve Carbon Neutrality Before 2060: Xi*. Accessed: Sep. 22, 2020. [Online]. Available: http://www.xinhuanet.com/english/2020-09/22/c_139388644.htm

[5] X. Liu and W. Xu, "Economic load dispatch constrained by wind power availability: A here-and-now approach," *IEEE Trans. Sustain. Energy*, vol. 1, no. 1, pp. 2–9, Apr. 2010.

[6] S. Impram, S. V. Nese, and B. Oral, "Challenges of renewable energy penetration on power system flexibility: A survey," *Energy Strategy Rev.*, vol. 31, Sep. 2020, Art. no. 100539, doi: 10.1016/j.esr.2020.100539.

[7] J. Lassila, J. Haakana, V. Tikka, and J. Partanen, "Methodology to analyze the economic effects of electric cars as energy storages," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 506–516, Mar. 2012, doi: 10.1109/tsg.2011.2168548.

[8] M. Ferdowsi, "Vehicle fleet as a distributed energy storage system for the power grid," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2009, pp. 2074–2075.

[9] R. Rana, M. Singh, and S. Mishra, "Design of modified droop controller for frequency support in microgrid using fleet of electric vehicles," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3627–3636, Sep. 2017, doi: 10.1109/tpwrs.2017.2651906.

[10] P. Jampeethong and S. Khomfoi, "Coordinated control of electric vehicles and renewable energy sources for frequency regulation in microgrids," *IEEE Access*, vol. 8, pp. 141967–141976, 2020, doi: 10.1109/access.2020.3010276.

[11] K. C. Divya, J. Ostergaard, E. Larsen, C. Kern, T. Wittmann, and M. Weinhold, "Integration of electric drive vehicles in the Danish electricity network with high wind power penetration," *Eur. Trans. Electr. Power*, vol. 20, no. 7, pp. 872–883, Oct. 2010, doi: 10.1002/etep.371.

[12] M. Gonzalez Vaya and G. Andersson, "Self scheduling of plug-in electric vehicle aggregator to provide balancing services for wind power," *IEEE Trans. Sustain. Energy*, vol. 7, no. 2, pp. 886–899, Apr. 2016, doi: 10.1109/tste.2015.2498521.

[13] W. Hu, C. Su, Z. Chen, and B. Bak-Jensen, "Optimal operation of plug-in electric vehicles in power systems with high wind power penetrations," *IEEE Trans. Sustain. Energy*, vol. 4, no. 3, pp. 577–585, Jul. 2013, doi: 10.1109/tste.2012.2229304.

[14] A. Alahyari, M. Ehsan, and M. Mousavizadeh, "A hybrid storage-wind virtual power plant (VPP) participation in the electricity markets: A self-scheduling optimization considering price, renewable generation, and electric vehicles uncertainties," (in English), *J. Energy Storage*, vol. 25, Oct. 2019, Art no. 100812, doi: 10.1016/j.est.2019.100812.

[15] M. Vasirani, R. Kota, R. L. G. Cavalcante, S. Ossowski, and N. R. Jennings, "An agent-based approach to virtual power plants of wind power generators and electric vehicles," *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1314–1322, Sep. 2013, doi: 10.1109/Tsg.2013.2259270.

[16] P. Kou, D. Liang, L. Gao, and F. Gao, "Stochastic coordination of plug-in electric vehicles and wind turbines in microgrid: A model predictive control approach," *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1537–1551, May 2016, doi: 10.1109/tsg.2015.2475316.

[17] Y. Zhu, H. Gao, J. Xiao, B. Qu, F. Zhu, and L. Yang, "Dynamic multi-objective dispatch considering wind power and electric vehicles with probabilistic characteristics," *IEEE Access*, vol. 7, pp. 185634–185653, 2019, doi: 10.1109/access.2019.2961242.

[18] J. Zhao, F. Wen, Z. Y. Dong, Y. Xue, and K. P. Wong, "Optimal dispatch of electric vehicles and wind power using enhanced particle swarm optimization," *IEEE Trans. Ind. Informat.*, vol. 8, no. 4, pp. 889–899, Nov. 2012, doi: 10.1109/tii.2012.2205398.

[19] C. D. Korkas, S. Baldi, S. Yuan, and E. B. Kosmatopoulos, "An adaptive learning-based approach for nearly optimal dynamic charging of electric vehicle fleets," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 7, pp. 2066–2075, Jul. 2018, doi: 10.1109/tits.2017.2737477.

[20] D. S. Kirschen, G. Strbac, P. Cumperayot, and D. de Paiva Mendes, "Factoring the elasticity of demand in electricity prices," *IEEE Trans. Power Syst.*, vol. 15, no. 2, pp. 612–617, May 2000, doi: 10.1109/59.867149.

[21] Z. Li and M. Ouyang, "The pricing of charging for electric vehicles in China—Dilemma and solution," *Energy*, vol. 36, no. 9, pp. 5765–5778, Sep. 2011, doi: 10.1016/j.energy.2011.05.046.

[22] Y. Zhuang, D. Yao, and Z. Zhao, "EV charging price mechanism," *East China Electr. Power*, vol. 42, no. 9, pp. 1938–1940, 2014.

[23] C. Wu, H. Mohsenian-Rad, and J. Huang, "Vehicle-to-aggregator interaction game," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 434–442, Mar. 2012, doi: 10.1109/tsg.2011.2166414.

[24] W. Tushar, W. Saad, H. V. Poor, and D. B. Smith, "Economics of electric vehicle charging: A game theoretic approach," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1767–1778, Dec. 2012, doi: 10.1109/tsg.2012.2211901.

[25] J. Yang, Y. J. Lin, F. Z. Wu, and L. Chen, "Subsidy and pricing model of electric vehicle sharing based on two-stage Stackelberg game—A case study in China," (in English), *Appl. Sci.-Basel*, vol. 9, no. 8, p. 1631, Apr. 2019, doi: 10.3390/app9081631.

[26] Q. Zhang, Y. Hu, W. Tan, C. Li, and Z. Ding, "Dynamic time-of-use pricing strategy for electric vehicle charging considering user satisfaction degree," *Appl. Sci.*, vol. 10, no. 9, p. 3247, May 2020, doi: 10.3390/app10093247.

[27] Q. Chen, F. Wang, B.-M. Hodge, J. Zhang, Z. Li, M. Shafie-Khah, and J. P. S. Catalao, "Dynamic price vector formation model-based automatic demand response strategy for PV-assisted EV charging stations," *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 2903–2915, Nov. 2017, doi: 10.1109/tsg.2017.2693121.

[28] Y. Nie, X. Wang, and K.-W.-E. Cheng, "Multi-area self-adaptive pricing control in smart city with EV user participation," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 7, pp. 2156–2164, Jul. 2018, doi: 10.1109/tits.2017.2759192.

[29] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015, doi: 10.1038/nature14236.

[30] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, Nov. 2017, doi: 10.1109/MSP.2017.2743240.

[31] T. Tongloy, S. Chuwongin, K. Jaksukam, C. Chousangsuntorn, and S. Boonsang, "Asynchronous deep reinforcement learning for the mobile robot navigation with supervised auxiliary tasks," in *Proc. 2nd Int. Conf. Robot. Autom. Eng. (ICRAE)*, Dec. 2017, pp. 68–72, doi: 10.1109/ICRAE.2017.8291355.

[32] G. Feng and A. Serletis, "Productivity trends in U.S. Manufacturing: Evidence from the NQ and AIM cost functions," *J. Econometrics*, vol. 142, no. 1, pp. 281–311, Jan. 2008, doi: 10.1016/j.jeconom.2007.06.002.

[33] J. Contreras, M. Klusch, and J. B. Krawczyk, "Numerical solutions to Nash–Cournot equilibria in coupled constraint electricity markets," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 195–206, Feb. 2004, doi: 10.1109/tpwrs.2003.820692.

[34] X. Zhao, S. Ding, Y. An, and W. Jia, "Applications of asynchronous deep reinforcement learning based on dynamic updating weights," *Int. J. Speech Technol.*, vol. 49, no. 2, pp. 581–591, Feb. 2019, doi: 10.1007/s10489-018-1296-x.

YANG PAN received the bachelor's degree from the School of Economics and Management, North China Electric Power University (NCEPU), in 2019, where she is currently pursuing the master's degree. Her main research interest includes electric vehicles.

WEIYE WANG received the bachelor's degree from the School of Economics and Management, North China Electric Power University (NCEPU), in 2019, where he is currently pursuing the master's degree. His main research interest includes electricity market.

YANBIN LI received the Ph.D. degree in management science and engineering from Beihang University, China. He is currently a Professor with the School of Economics and Management, North China Electric Power University (NCEPU), China. His research interest includes electric power enterprise development management.

FENG ZHANG received the bachelor's degree from the School of Economics and Management, North China Electric Power University (NCEPU), in 2016, where he is currently pursuing the Ph.D. degree. His main research interest includes energy management.

YANTING SUN received the bachelor's degree from the School of Economics and Management, North China Electric Power University (NCEPU), in 2017, where she is currently pursuing the Ph.D. degree. Her main research interest includes supply chain management.

DUNNAN LIU received the B.E. and Ph.D. degrees from Tsinghua University, China, both in electrical engineering. He is currently an Associate Professor with the School of Economics and Management, North China Electric Power University (NCEPU), China. His research interests include risk management and operation of power market.