IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 28, 2021, accepted June 16, 2021, date of publication June 23, 2021, date of current version July 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3091725

Strategy of Large-Scale Electric Vehicles Absorbing Renewable Energy Abandoned Electricity Based on Master-Slave Game

DUNNAN LIU, LINGXIANG WANG, WEIYE WANG, HUA LI, MINGGUANG LIU, AND XIAOFENG XU

School of Economics and Management, North China Electric Power University, Beijing 102206, China
Beijing Key Laboratory of New Energy and Low Carbon Development, North China Electric Power University, Beijing 102206, China

Corresponding author: Lingxiang Wang (120192206112@ncepu.edu.cn)

This work was supported in part by the National Social Science Foundation of China under Grant 19ZDA081.

ABSTRACT The rapid development of renewable energy power has improved global energy and environmental problems. However, with the high volatility of renewable energy, it is an important challenge to guarantee the consumption of renewable energy and the reliable operation of high percentage renewable energy power systems. To solve this problem, this paper proposes a tracking absorption strategy for renewable energy based on the interaction between the supply side and the demand side, which adjusts the charging process of electric vehicles (EVs) through electric vehicle aggregator (EVA) to realize the tracking absorption of renewable energy abandoned electricity. In view of this process, we analyze the interaction among power grid, EVA and renewable energy generation (REG) as well as their market characteristics. The master-slave game model of EVA and REG was constructed considering the charging behavior characteristics of EVs and the output characteristics of REGs. Then the model solving strategy based on soft actor-critic (SAC) algorithm is proposed, and the REG pricing strategy and EVA scheduling strategy are calculated to optimize the mutual benefits. The case analysis shows that, under the same scale of electric vehicles, the proposed method can promote about 93.89% of the power abandonment consumption of wind power system, 96.00% of the photovoltaic system, and 97.41% of the wind-solar system. This strategy reduces the electricity purchase cost of EVA, promotes the interaction among renewable energy, vehicles and power grid, and improves the utilization efficiency of renewable energy.

INDEX TERMS Reinforcement learning, electric vehicles, renewable energy abandoned power consumption, curve tracking, V2G.

I. INTRODUCTION

China's renewable energy is in a period of rapid development, by the end of 2020, the total installed capacity of renewable energy generation in China reached 930 million kW, accounting for 42.4% of the total installed capacity, and the National Energy Administration expects that by the 14th Five-Year Plan period, the proportion of clean energy in the incremental energy consumption will reach 80%. The consumption of renewable energy will be one of the key issues facing the development of renewable energy in China, and the aggregated dispatch of EVs can realize the consumption of renewable energy. By the end of 2020, the number of EVs in China will reach 4.7 million, and the number of charging piles will reach 1.148 million, which has the potential to aggregate and consume renewable energy.

Because the renewable energy has great instability, in the case of sufficient power supply, the renewable energy power is often discarded to choose power resources to ensure the economic and reliable operation of the power system [1]. In response to this situation, there are two main measures, one is to improve the prediction accuracy of renewable energy sources, and the other is to deploy energy storage or other coupling to suppress load fluctuations in the form of energy transfer [2]–[4]. For example, Zhou *et al.* [5] proposed a wind speed forecasting method with heteroscedastic

The associate editor coordinating the review of this manuscript and approving it for publication was Ruisheng Diao.

multi-kernel learning controlling the wind power forecasting error between 4.57% and 8.751%, helpfully adjusts the unit scheduling plan in time. This method can reduce the power abandonment loss, but it cannot effectively eliminate the power abandonment loss. Wang *et al.* [6] studied the optimal ratio of photovoltaic and battery. The battery is used to store and release electric energy to ensure power supply, and the discarded electricity can be reduced from 100% to 0% by putting into the energy storage battery. However, this method needs to invest energy storage resources and pay a large cost. Through the price incentive of renewable energy generators and the response scheduling of Electric Vehicle Aggregator (EVA), the consumption of renewable energy can be promoted while controlling the cost.

Renewable energy output data varies on different days and at different time periods, and EVs show different adjustable capabilities at different incentive levels. Therefore, regulating the charging and discharging process of EVs by means of real-time price guidance is a reliable approach. In a study by Luo *et al.* [7], a dynamic price mechanism was proposed to regulate the subject behavior, which increased the profit of renewable energy power plants by 7% and adjusted the charging load of EVAs. However, the game process between the generation side and the user side in the price setting process should be considered. This article has made a special design to this point.

How to design an optimal dispatching strategy for large-scale electric vehicles is an important issue for EV to absorb renewable energy. The current research on the optimal dispatching process for large-scale EVs can be summarized in two points [8]–[10]: 1) How to motivate adjustable EVs to participate in the electricity market? To address this problem, Gan *et al.* [11] studied charging behavior of EV owners and designed a corresponding price incentive method to realize load transfer through price guidance. However, user behavior is an individual characteristic, and its sample data is difficult to collect [12], [13], its typicality is questionable, and it is not universal [14]. Therefore, we transfer the study of user behavior to EVA, to achieve orderly charging of electric vehicles. EVA acts as the agent for users to respond to price signals of renewable energy and EVA selects EVs willing to participate in the scheduling to respond to load regulation. 2) How to design a dispatch plan for large-scale EVs? In current research, optimal scheduling plans can be formed by solving optimization algorithms, including non-cooperative game [15], Probabilistic, coordinated charging method [16] and other methods. However, these methods are limited by their algorithm complexity, computational scale, and other conditions, and cannot solve the real-time scheduling problem of large-scale EVs. However the soft actor-critic algorithm based on deep reinforcement learning (DPL) mechanism can achieve large-scale and fine scheduling of EVs [17]. The SAC algorithm overcomes the defects of traditional algorithms and can provide support for real-time EVs scheduling.

For the first time, this paper studies the method of consuming renewable energy on the demand side, and proposes a strategy of reasonably scheduling EVS charging process to absorb renewable energy power. EVA reduces the charging cost of EVs by purchasing abandoned electricity, and REG increases the revenue of electricity sales by increasing electricity sales. The whole process promotes the consumption of renewable energy in an interactive way between the supply side and demand side of electricity, and maintains the economy, cleanliness and reliability of the power system operation. In general, this paper proposes a market model for EVA to track and consume abandoned electricity; for the market game process of EVA and REG, it proposes a master-slave game based REG pricing model and EVA dispatching model, which is helpful to achieve a win-win situation for multiple subjects; and it proposes a model solving method based on SAC deep reinforcement learning, which realizes real-time optimal dispatching of large-scale electric vehicles; the interaction between wind power, photovoltaic, wind-solar systems and EVA is analyzed, and the energy exchange efficiency between different energy systems and EVA is studied.

The first part of this paper is the introduction, which discusses the development of EV and renewable energy; the other parts of this paper are structured as follows: section 2 dissects the scenario of EVA tracking renewable energy for abandoned power consumption; section 3 constructs the behavioral model and market game model of EVA tracking REG absorbing abandoned power; section 4 establishes the model solving algorithm based on SAC deep reinforcement learning; section 5 constructs a case study of EVA tracking for renewable energy consumption which verifies the effectiveness of this paper's model; section 6 concludes and discusses the whole paper.

II. FRAMEWORK OF THE SCENE

A. SCHEDULING FRAMEWORK FOR EV TRACKING RENEWABLE ENERGY CONSUMPTION

In a conventional power system, REG acts as an important generation resource to deliver power to the grid and provide energy supply to demand-side users; EVA acts as an aggregator to purchase power from the electricity market for EVs and obtains power through the grid thus charging EVs [18], [19]. In the process of delivering power to the grid by the REG, when the power delivery channel is full or the power demand is satisfied, the REG has to discard some of the power. In the process of EVA receiving power from the grid, the charging time, the power of charging, and the charging quantity of EV can be adjusted, but they are not utilized.

Therefore, this paper proposes a power tracking and consumption strategy to reduce the power purchase cost of EVAs and promote the consumption of renewable abandoned power through direct trading between EVs and REGs. The response strategy of EVs tracking and consuming renewable energy is shown in Figure 1.

The process promotes the consumption of renewable energy and reduces the charging cost of EVAs through direct electricity trading between REG and EVAs. The process has the following characteristics.

(1) Most of RGE's electricity is sold through the wholesale market and transmitted through the grid to customers across the country. Feed-in electricity of RGE depends on the grid electricity demand, and the feed-in price is determined by RGE's market bidding in the wholesale market. Generation resources are mainly considered wind power and photovoltaic (PV).

(2) EVA receives most of its electricity from grid transmission. Moreover, EVA accepts electricity from state grid at wholesale market prices.

(3) Part of RGE's electricity is abandoned because the power grid cannot fully accept it. For this part of electricity, RGE directly trades with EVA to achieve renewable energy tracking and consuming, which is a kind of small-scale vehicle to grid (V2G). And in the process, RGE sends the price signal to EVA; then, EVA schedules the charging and discharging process of EVs in real time according to its own power demand and power purchase cost; finally, it completes the adjustment of load distribution and consumes renewable energy power.

B. SCHEDULING PROCESS

In this paper, the process of EVs tracking renewable energy is as follows: First, RGE makes a short-term forecast based on the generation of the units, and compares the forecasted power output with the current power contract to get the power abandonment curve. Then, RGE sends a transaction request to EVA for power tracking and consumption, along with a time-divided power trading volume quota and price demand. After receiving the trading request from RGE, EVA makes a reasonable power dispatch on the same day, taking into account the power system constraints and economic constraints, and trying to consume this part of the abandoned power. Finally, EVA and RGE checked the renewable energy power tracked and consumed, and settle the amount of electricity according to the time-divided price published by RGE.

III. MODEL FOR TRACKING THE CONSUMPTION OF RENEWABLE ENERGY

A. ELECTRIC VEHICLE BEHAVIOR MODEL

According to the charging and discharging process of EV, the single EV behavior model is constructed as shown in equation (1)-equation (6). Where equation (1) is the EV electric energy description formula; equation (2) describes the charging satisfaction situation when the EV leaves; equation (3) is the EV electric energy change constraint; equation (4)-equation (6) is the EV charging power constraint [20].

$$d_{ev,t} = \eta \sum_{k=t_{in}}^{t} P_{ev,k}\Delta t = d_{t-1} + \eta P_{ev,t}\Delta t \tag{1}$$

$$d_{min,t_{out}} = d_{max,t_{out}} = D_{expect} \tag{2}$$

$$d_{min,t} \leq d_{ev,t} \leq d_{max,t} \tag{3}$$

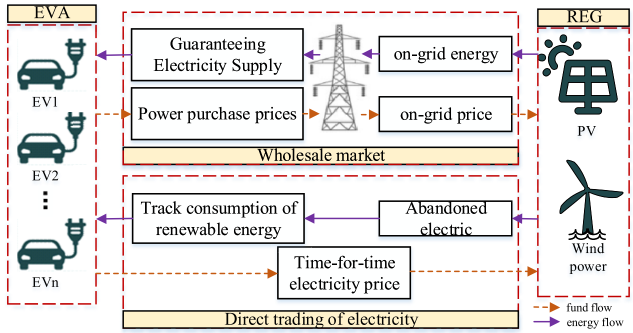

FIGURE 1. The response strategy of EVs tracking and consumption of renewable energy.

$$P_{max,t} = \frac{\min\left(P_{ev,max}, d_{max,t} - d_{t-1}\right)}{\eta\Delta t} \tag{4}$$

$$P_{min,t} = \frac{\max\left(P_{ev,min}, d_{min,t} - d_{t-1}\right)}{\eta\Delta t} \tag{5}$$

$$P_{min,t} \leq P_{ev,t} \leq P_{max,t} \tag{6}$$

where, $d_{ev,t}$ is the energy trajectory value of EV at the time of t; $d_{min,t_{out}}$ and $d_{max,t_{out}}$ are the lower limit and upper limit of the energy trajectory d_t at time t; η is charging efficiency; $P_{ev,k}$ is the constant power in the period from time k to time $k+1$; Δt is the time interval of the scheduling period; t_{in} and t_{out} are the time of EV access and departure, that is, EV access at the time of t_{in} and departure at the time of t_{out}; D_{expect} is charging demand of EV; $P_{ev,max}$ is the upper limit of rated charging power of EV battery; $P_{ev,t}$, $P_{max,t}$ and $P_{min,t}$ are the charging power of EV and the maximum and minimum charging power restricted by energy boundary constraints at the time t.

At the same time, for the aggregation process of EVs, there are equations (7)–(13). Among them, equation (7) is the calculation formula of EVA electrical energy; equation (7)-equation (10) is the constraint of the electric quantity of EVA purchase; equation (11)-equation (13) is the constraint of EVA load.

$$d_{m,t} = \eta \sum_{k=1}^{t} P_{m,k,t}\Delta t = d_{m,t-1} + \eta P_{m,t}\Delta t \tag{7}$$

$$d_{min,m,t} \leq d_{m,t} \leq d_{max,m,t} \tag{8}$$

$$d_{min,m,t} = \sum_{l=1}^{n_m} d_{min,m,l} \tag{9}$$

$$d_{max,m,t} = \sum_{l=1}^{n_m} d_{max,m,l} \tag{10}$$

$$P_{m,t} = \sum_{l=1}^{n_m} P_{ev,m,l,t} \tag{11}$$

$$P_{min,m,t} = \sum_{l=1}^{n_m} P_{ev,min,m,l,t} \leq P_{m,t} \tag{12}$$

$$P_{max,m,t} = \sum_{l=1}^{n_m} P_{ev,min,m,l,t} \geq P_{m,t} \tag{13}$$

where, $d_{m,t}$ is the energy trajectory of the subset group at time t; $d_{min,m,t}$ and $d_{max,m,t}$ are the upper and lower limits of the energy trajectories of the subset group at time t; $P_{m,t}$ is the total charging power of the subset group at time t; $P_{min,m,t}$ and $P_{max,m,t}$ are respectively the charging lower limit and upper limit of the subset group at time t; n_m is the total number of EVs belonging to EVA at time t; $d_{min,m,l,t}$, $d_{max,m,l,t}$ and $P_{ev,min,m,l,t}$, $P_{ev,min,m,l,t}$ are respectively the lower and upper bounds of the energy trajectory of the EV of the subset group l at time t and the lower and upper bounds of the charging power.

B. TRACKING AND ABSORBING MODEL BASED ON MASTER-SLAVE GAME

In the process of tracking the consumption, firstly, the REG puts forward the load regulation demand based on the predicted output value, and then the EVA will carry out the load regulation. In this process, REG is in the leading position as the main body of the consumption curve, and EVA is in the following position as the executor of the load consumption curve. In this process, REG will release the load consumption curve and its corresponding consumption price, and EVA will make consumption behaviors in response to the curve and consumption prices and create feedback on the pricing strategy of REG. And finally achieve a dynamic balance between the REG and the EVA, forming a stable price level [21]–[23].

This master-slave game can be described as follows.

$$\max \sum_{t=1}^{T} \left(P_{load,t} \cdot p_t + P_{ev2,t} \cdot p_{ev,t}\right) \tag{14}$$

S.t. (1)–(13)

$$P_{ev2,t} \le P_{wt,t} + P_{pv,t} - P_{load,t} \tag{15}$$

$$p_{ev,t} < p_t \tag{16}$$

where, $P_{load,t}$ is the total load at time t; T is the total number of periods; p_t is the electricity selling price at time t; $P_{ev2,t}$ is the amount of renewable energy absorbed by EVA at a moment; $p_{ev,t}$ is the time-divided electricity price of renewable energy power purchased by EVA at time t.

In the master slave game, the renewable energy power plant establishes the optimal consumption electricity price through the optimized Equations (14)–(16), and transmits the electricity price to EVA through communication equipment. Then EVA formulates charging strategy P_{ev} with the lowest electricity purchase cost as the optimization goal, and its model is as follows.

$$P_{ev} = \arg\min \sum_{t=1}^{T} \left(P_{ev1,t} \cdot p_t + P_{ev2,t} \cdot p_{ev,t}\right) \tag{17}$$

$$\sum_{t=1}^{T} \left(P_{ev1,t} \cdot p_t + P_{ev2,t} \cdot p_{ev,t}\right) = \sum_{t=1}^{T} P_{m,t} \tag{18}$$

$$P_{ev1,t} + P_{ev2,t} = P_{m,t} \tag{19}$$

where, $P_{ev1,t}$ is the electric quantity directly purchased by EVA at time t.

IV. SAC BASED MODEL SOLVING ALGORITHM

In contrast to traditional distributed optimization algorithms, reinforcement learning (RL) allows continuous interaction with the environment through the "trial and error" process of the agent. It uses a model-based reward mechanism to seek solutions that maximize cumulative benefits, and is an optimization decision approach for the interaction of EVs with sources, networks, and loads [24]. The SAC algorithm is a reinforcement learning algorithm proposed by T. Haarnoja *et al.* [25]. It can make accurate and effective charging and discharging decisions for large-scale EVs in complex power supply and demand environments by introducing a maximum entropy encouragement strategy to improve the robustness of the algorithm while accelerating the training speed [25], [26].

A. IMPORTANT CONCEPTS

SAC algorithm, by introducing entropy into RL [27], [28], makes the policy as random as possible and then enables agents to explore the policy space more fully. Its important concepts include policy, entropy, and soft value function.

1) THE MAXIMUM REWARD POLICY

$$\pi^*_{max} = \arg\max_{\pi} \sum_{t=1}^{T} \mathbb{E}_{(s_t,a_t)\sim p_t}\left[r\left(s_t, a_t\right) + \alpha\mathcal{H}\left(\pi\left(\cdot|\, s_t\right)\right)\right] \tag{20}$$

where, s_t is the state space of EVA at time t (i.e., system base load); a_t is the action space of EVA at time t (i.e., the charge and discharge conditions); $r(s_t, a_t)$ is the reward function of EVA at time t, which is reflected in equation (20) in this paper; $(s_t, a_t) \sim p_\pi$ is the state-action trajectory formed by strategy π; p_π represents the distribution followed by the state-action pair that the agent will encounter under the policy π; α is temperature, which determines the effect of entropy on the reward; $\mathcal{H}(\pi(\cdot|\, s_t))$ is the entropy of the strategy in state s_t, and its calculation method is shown in equation (21).

2) ENTROPY

It is used to measure the randomness of a random variable, and the random distribution it follows is directly considered in the actual calculation.

$$\begin{aligned}\mathcal{H}\left(\pi\left(a_t|\, s_t\right)\right) &= -\int_{a_t} \pi\left(a_t|\, s_t\right) \log\left(\pi\left(a_t|\, s_t\right)\right) da_t \\ &= \mathbb{E}_{a_t\sim p_\pi}\left[-\log\left(\pi\left(a_t|\, s_t\right)\right)\right]\end{aligned} \tag{21}$$

3) SOFT VALUE FUNCTION

Soft Value Function can be used to evaluate the goodness of the strategy. The soft Q function can be calculated as in

Equation (22).

$$Q\left(s_q, a_q\right) = r\left(s_q, a_q\right) + \gamma \mathbb{E}_{s_{t+1}\sim p}\left[Q\left(s_{q+1}, a_{q+1}\right)\right] \quad (22)$$

where, γ is the discount factor of reward.

And the calculation formula of Soft Value Function is shown in (23).

$$V\left(s_t\right) = \mathbb{E}_{a_t\sim\pi}\left[Q\left(s_t, a_t\right) - \log \pi\left(a_t \mid s_{t+1}\right)\right] \quad (23)$$

B. POLICY ITERATION

1) SOFT POLICY EVALUATION

Fixed policy, using Bellman equation (24) to update Q value until convergence.

$$Q\left(s_t, a_t\right) = r\left(s_t, a_t\right) + \gamma \mathbb{E}_{s_{t+1}\sim p}\left[Q\left(s_{t+1}, a_{t+1}\right) - \alpha \log \pi\left(a_{t+1} \mid s_{t+1}\right)\right] \quad (24)$$

The convergence can be achieved by iteration based on Equations (22)-(23).

2) SOFT POLICY IMPROVEMENT

$$\pi'\left(s\right) = \arg\min_{\pi_k\in\prod} D_{\mathrm{KL}}\left(\pi_k\left(\cdot \mid s_q\right) \middle\| \frac{\exp\left(\frac{1}{\alpha}Q^{\pi_{\mathrm{old}}}\left(s_k, \cdot\right)\right)}{Z^{\pi_{\mathrm{old}}}\left(s_k\right)}\right) \quad (25)$$

where, D_{KL} is KL divergence; $\prod$ is the policy set; $Z^{\pi_{old}}\left(s_q\cdot\right)$ is the partition function, and is used for the normalized distribution.

C. SOFT ACTOR-CRITIC

First, we need to build a Q value network and a policy network. The Q-value network outputs single-value Q through several layers of neural networks, and the policy network outputs a Gaussian distribution [29]. In this process, the neural network will be updated. The Q value network parameter has an update strategy as shown in Formula (26), and the policy network parameter has an update strategy as shown in Formula (27).

$$J_Q\left(\theta\right) = \mathbb{E}_{\left(s_t, a_t, s_{t+1}\right)\sim\mathcal{D}}\left[\frac{1}{2}\left(Q\left(s_t, a_t\right) - \left(r\left(s_t, a_t\right) + \gamma V_{\bar{\theta}}\left(s_{t+1}\right)\right)\right)^2\right] \quad (26)$$

$$J_\pi\left(\phi\right) = D_{\mathrm{KL}}\left(\pi\left(\cdot \mid s_t\right) \middle\| \exp\left(\frac{1}{\alpha}Q_\theta\left(s_t, \cdot\right) - \log Z\left(s_t\right)\right)\right) \quad (27)$$

where, θ is the Q value network parameter.; ϕ is the policy network parameter.

What's more, we usually give a fixed temperature as the weight of entropy. But in fact, due to the constant changes of rewards, it is not reasonable to use a fixed temperature, which will make the whole training unstable. Therefore, a constrained optimization method (28)-(29) is considered so that the mean value of temperature is limited and variable under different states.

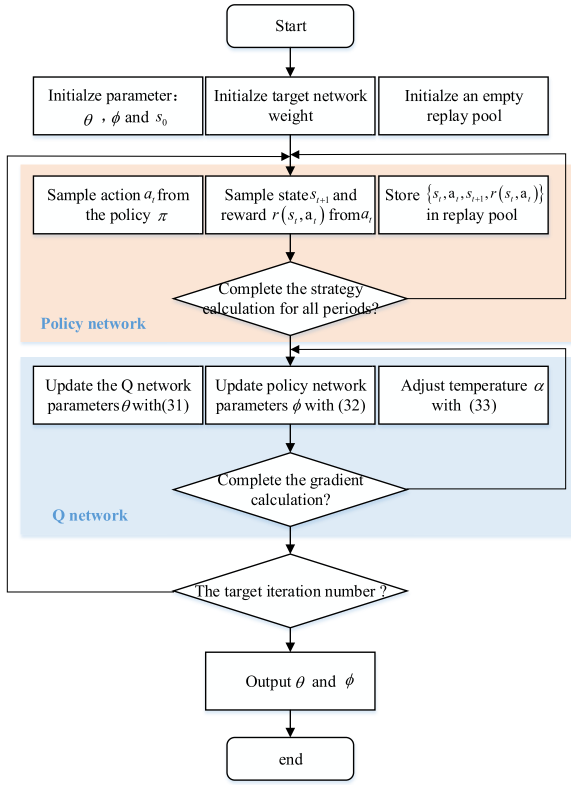

FIGURE 2. Calculation flow of EVA optimization scheduling based on SAC.

$$\max_{\pi_k\in\prod} \mathbb{E}\left[\sum_{t=0}^{T} r\left(s_t, a_t\right)\right] \quad (28)$$

$$s.t.\ \forall\mathcal{H}\left(\pi_t\right) \geq \mathcal{H}_0 \quad (29)$$

After solving, we can get the loss of temperature as equation (30).

$$J\left(\alpha\right) = \mathbb{E}_{a_t\sim\pi_t}\left[-\alpha \log \pi_t\left(a_t \mid \pi_t\right) - \alpha\mathcal{H}_0\right] \quad (30)$$

D. CALCULATION FLOW

In the process of EVA tracking and absorbing renewable energy abandoned electricity, a master-slave game is adopted, which fully considers the scale characteristics and market positions of both sides. SAC algorithm is used to solve the problem, which fully considers the complexity characteristics of calculation scale and decision. The calculation flow of EVA optimization scheduling based on SAC is shown in Figure 2.

V. CASE STUDY

A. BASIC SITUATION OF THE CASE

This paper takes the output data and load data of a typical scene in Northeast China as an example to analyze. In this case, the typical output curves of wind power generation in different seasons are shown in Fig. 3, the typical output situation of photovoltaic power generation in different seasons is

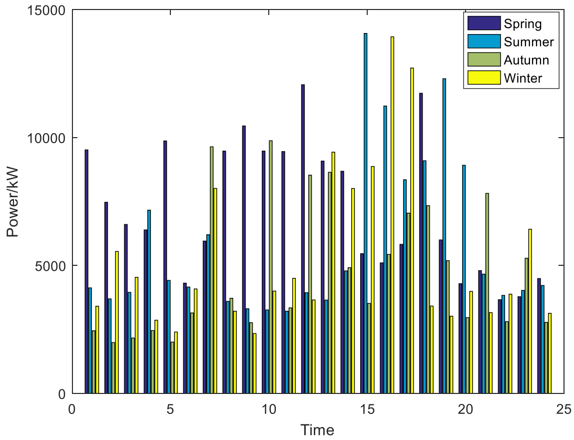

FIGURE 3. Wind power generation.

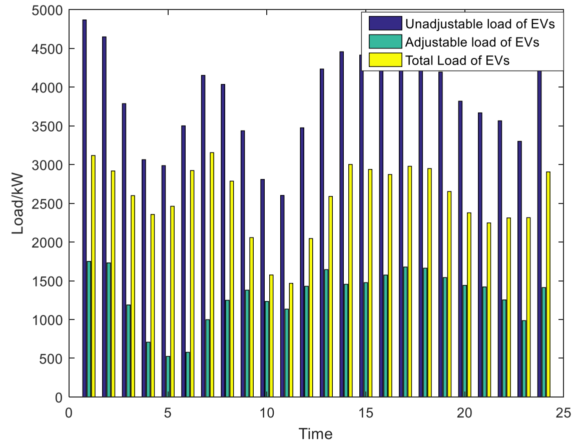

FIGURE 6. Load situation of EVs.

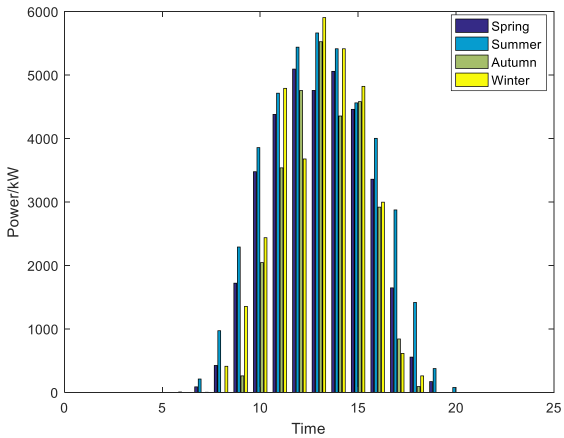

FIGURE 4. Photovoltaic power generation.

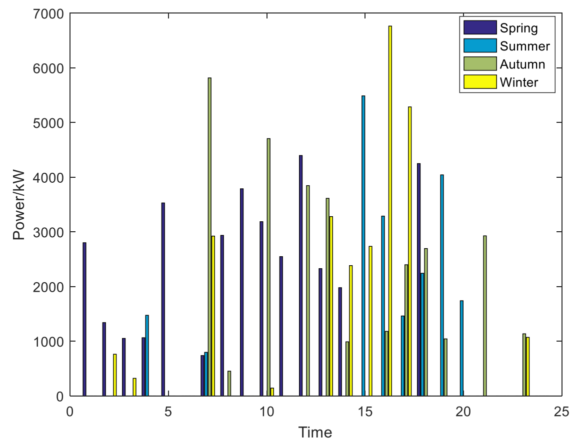

FIGURE 7. Power to be consumed in scenario 1.

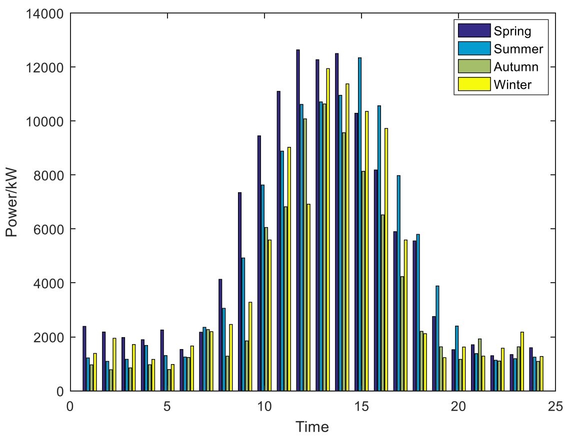

FIGURE 5. Power generation of wind-solar system.

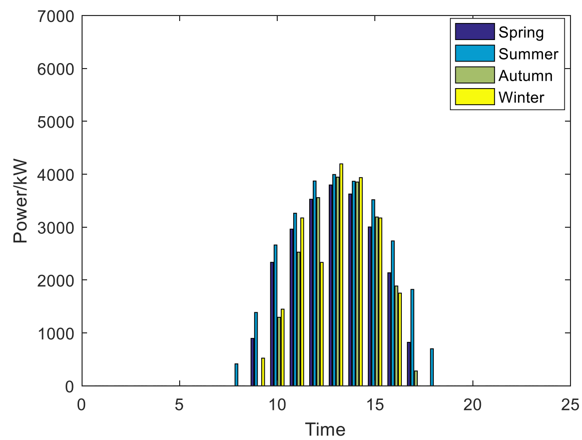

FIGURE 8. Power to be consumed in scenario 2.

shown in Figure 4, the typical output situation of wind-solar system in different seasons is shown in Figure 5, and the load of EVs is shown in Figure 6.

In this section, 3 typical scenarios are designed for analysis. Scenario 1 is the renewable energy tracking and consumption of wind power; Scenario 2 documents the renewable energy tracking and consumption of PV; Scenario 3 is the renewable energy tracking and consumption of wind power and PV. Figure 7, Figure 8 and Figure 9 show the distribution of abandoned electricity from renewable energy under three scenarios.

The on-grid electricity price of wind power in this region is 0.57yuan/kWh, including subsidies of 0.179yuan/kWh; The on-grid electricity price of photovoltaic is 0.75yuan/kWh, including subsidy 0.37yuan/kWh; The electricity price for users to purchase electricity from the grid side is 0.64yuan/kWh. All data are the result of local real data processing in 2019.

TABLE 1. The consumption of renewable energy abandoned electricity.

Scenarios	Label	Spring	Summer	Autumn	Winter	Summary values
	Initial abandoned power (kWh)	35932.48	20532.85	30799.27	25666.06	112930.66
Scenario 1	EV tracks power consumption(kWh)	34367.41	19860.19	27747.28	24051.00	106025.88
	Percentage of consumption	95.64%	96.72%	90.09%	93.71%	93.89%
	Initial abandoned power(kWh)	23099.45	28232.66	20532.85	20532.85	92397.81
Scenario 2	EV tracks power consumption(kWh)	22214.88	27499.74	19044.11	19943.41	88702.14
	Percentage of consumption	96.17%	97.40%	92.75%	97.13%	96.00%
	Initial abandoned power (kWh)	28432.66	28532.96	20432.85	20532.85	97931.32
Scenario 3	EV tracks power consumption(kWh)	27140.95	27658.28	20214.50	20379.32	95393.05
	Percentage of consumption	95.46%	96.93%	98.93%	99.25%	97.41%

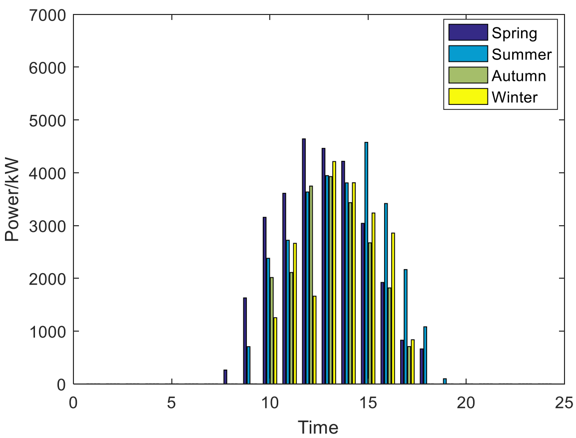

FIGURE 9. Power to be consumed in scenario 3.

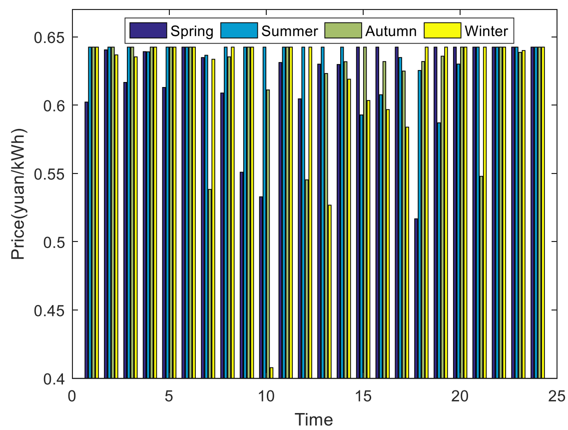

FIGURE 10. Time-divided price of REG for scenario 1.

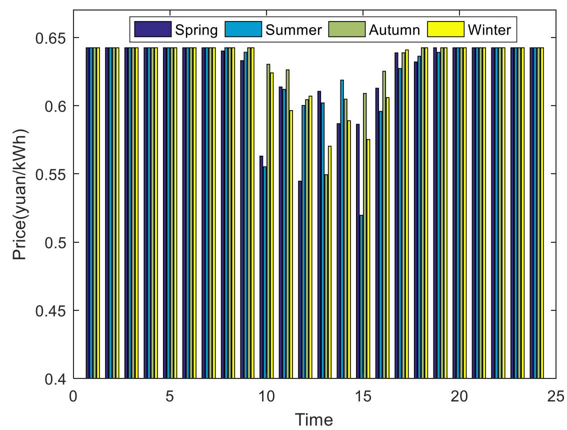

FIGURE 11. Time-divided price of REG for scenario 2.

In this section, we solve the pricing strategy of REG and the optimal scheduling strategy of EVA. Then we calculate the increased revenue and the increased consumption of renewable energy under three scenarios, and prove the economy and effectiveness of the model and method proposed in this paper. Finally, by comparing the optimal scheduling results of the three scenarios, the coupling performance between different energy sources and EVs can be judged, which provides a reference for EVA to select a suitable cooperative manufacturer.

The computing environment is Python TensorFlow 2.0. The computer model is a quad-core 2.60-GHz Intel Core i7-6700HQ processor with 16GB of RAM.

B. CALCULATION RESULTS

The reward function of SAC algorithm is constructed based on equation (14)-(19), the SAC iterative network is constructed based on equation (20)-(30), and the solution constraints are generated based on equation (1)-(13). After 3000 iterations of solving and 8 hours of calculation, the optimal time-divided price of REG for scenario 1 is obtained as shown in Figure 10, the optimal time-divided price of REG for scenario 2 is obtained as shown in Figure 11, the optimal time-divided price of REG for scenario 3 is obtained as shown in Figure 12. The optimal EVA optimized scheduling strategy for scenario 1 is shown in Figure 13, and the optimal EVA optimized scheduling strategy for scenario 2 is shown in Figure 14, and the optimal EVA optimized scheduling strategy for Scenario 3 is shown in Figure 15.

In order to compare the load absorbing capacity of EV in three typical scenarios, the consumption of renewable energy abandoned electricity is calculated as shown in Table 1, and the revenue growth of REG and EVA is calculated as shown in Table 2.

TABLE 2. The revenue growth of REG and EVA.

Scenarios	Label	Spring	Summer	Autumn	Winter
	Incremental return of REG (yuan)	20408.87	12087.33	16511.15	14281.32
Scenario 1	Incremental return of EVA (yuan)	1648.14	658.94	1297.06	1154.60
	Total incremental revenue (yuan)	22057.01	12746.27	17808.2	15435.93
	Incremental return of REG (yuan)	13025.94	16387.16	11050.81	11936.61
Scenario 2	Incremental return of EVA (yuan)	1231.57	1262.17	1171.70	863.07
	Total incremental revenue (yuan)	14257.51	17649.33	12222.51	12799.68
	Incremental return of REG (yuan)	16072.04	16378.54	12167.15	12067.86
Scenario 3	Incremental return of EVA (yuan)	1347.02	1372.55	806.51	1011.58
	Total incremental revenue (yuan)	17419.06	17751.09	12973.66	13079.45

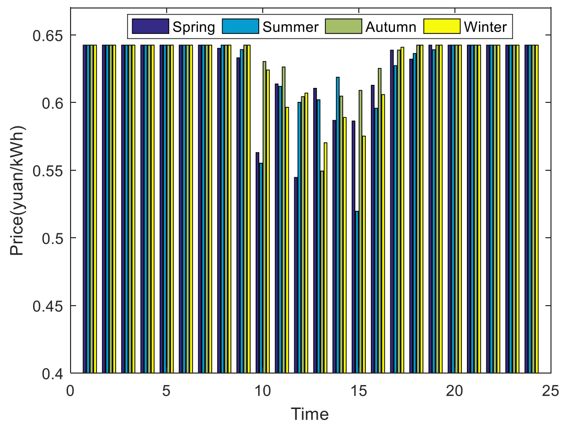

FIGURE 12. Time-divided price of REG for scenario 3.

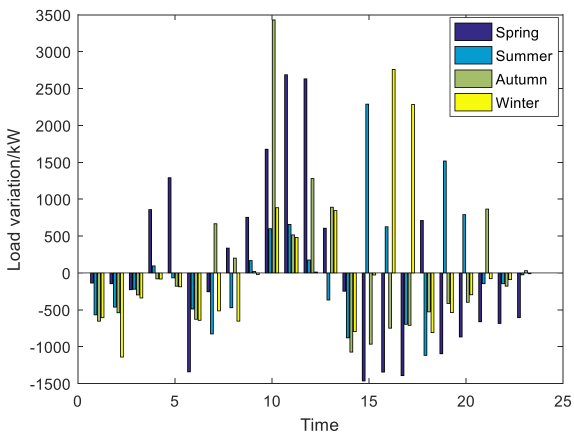

FIGURE 13. EVA optimized scheduling strategy for scenario 1.

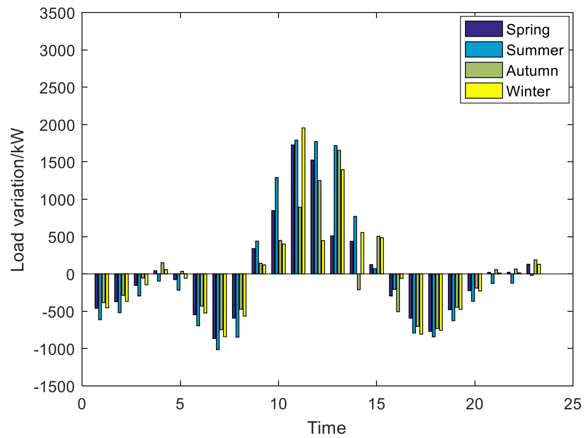

FIGURE 14. EVA optimized scheduling strategy for scenario 2.

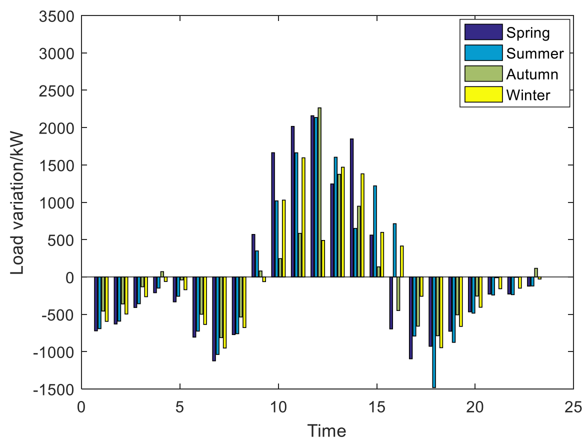

FIGURE 15. EVA optimized scheduling strategy for scenario 3.

Table 1 illustrates EVA's tracking and consumption of renewable energy in different seasons and different scenarios. Comparing the consumption in different seasons, it can be found that wind resources are more abundant in spring and PV resources are more abundant in summer. And in the season of more abundant resources, the proportion of EVA tracking consumption of renewable energy will not weaken. This is because even in the season with a lot of electric abandoning, the electric abandoning does not produce new peak value, but the time distribution is more extensive. Therefore, EVA can still absorb this part of electric abandoning by fully scheduling the charging process of EV.

Comparing the renewable energy consumption in different scenarios, we can find that the highest proportion of renewable energy is tracked and consumed in scenario 3, followed by scenario 2 and finally scenario 1. It can be seen that for wind power system, PV system and wind-solar system, the coupling of wind-solar system promotes the tracking and consumption of renewable energy by EVA; the more regular abandonment distribution of PV system is also conducive to the tracking and consumption of renewable energy by EVA. While the discrete and spike characteristics of wind

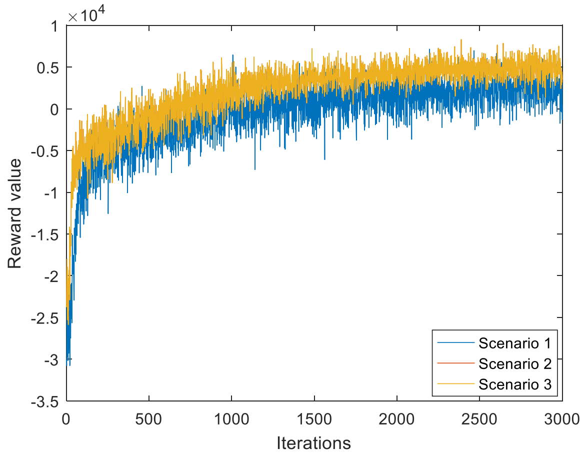

FIGURE 16. **Algorithm convergence graph.**

power system are not conducive to EVA's high proportion of consumption of renewable energy.

Table 2 compares the incremental benefits of each market player after the absorption of renewable energy power by EVA in different scenarios. The larger the scale of power abandonment, the higher the incremental benefit generated by EVA tracking absorption, such as spring in scenario 1, summer in scenario 2, and summer in scenario 3. In the three typical scenarios, REG increases its revenue by selling the abandoned power at an average price of 0.59 yuan/kWh; EVA saves the power purchase cost of 0.05 yuan/kWh on average.

Figure 16 is the SAC algorithm convergence graph, which shows the convergence of the SAC algorithm during the training process.

In the case of real-time updating of REG pricing policy and EVA scheduling policy, the SAC algorithm shows some oscillation in the convergence process due to the irregularity of the data source. However, it began to realize positive revenue and steady increase after 200 iterations, and achieved stable test results and good overall performance after 1300 iterations. SAC algorithm is a reinforcement learning algorithm suitable for EVA tracking and absorbing renewable energy.

VI. CONCLUSION

In order to promote the consumption of renewable energy and the development of V2G, this paper uses EVA to aggregate and schedule the charging process of EVS, realizes the tracking absorption of renewable energy's abandoned power, and solves the problem of power supply and demand balance under the condition of large-scale renewable energy input into the network. In the market response process, REG sends price signals and EVA accepts and responds to them in a way that ensures supply-side revenue and controls recipient cost. The scheduling process is carried out by the EVA, taking its own power purchase cost and customers' charging demands into account, to make reasonable and optimal scheduling. At the same time, a master-slave game model between REG and EVA is constructed to form a pricing strategy for REG and a scheduling strategy for EVA to maximize the interests of both parties by combining the economic and physical constraints of the system. The SAC deep reinforcement learning algorithm is used to solve the problems of high data dimension and large sample dispersion in the model solution process.

The model in this paper is verified based on the basic situation of A region. The results show that the model in this paper can achieve 93.89% of the power consumption of wind power system, 96.00% of the photovoltaic system, and 97.41% of the wind-solar system. Created a daily economic return of not less than 10,000 yuan (about 25%) for REG, and a daily economic return of not less than 1,000 yuan (about 7%) for EVA. The results prove the effectiveness and economy of the strategy in this paper, and have important reference value for promoting the consumption of renewable energy and the development of electric vehicles. The market transaction organization method between wind power, photovoltaic and EVA is the next research content.

REFERENCES

[1] J. Sun, M. Li, Z. Zhang, T. Xu, J. He, H. Wang, and G. Li, "Renewable energy transmission by HVDC across the continent: System challenges and opportunities," *CSEE J. Power Energy Syst.*, vol. 3, no. 4, pp. 353–364, Dec. 2017, doi: 10.17775/cseejpes.2017.01200.

[2] Y. Chen, R. Chen, Y. Shen, P. Zhang, Z. Liu, Y. Lang, and X. Yang, "A novel robust state estimator based on rectangular pulse function," *IEEE Access*, vol. 7, pp. 113525–113533, 2019, doi: 10.1109/ACCESS.2019.2931935.

[3] Y. Chen, Y. Yao, and Y. Zhang, "A robust state estimation method based on SOCP for integrated electricity-heat system," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 810–820, Jan. 2021, doi: 10.1109/TSG.2020.3022563.

[4] H. Jia, Y. Ding, Y. Song, C. Singh, and M. Li, "Operating reliability evaluation of power systems considering flexible reserve provider in demand side," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3452–3464, May 2019, doi: 10.1109/TSG.2018.2827670.

[5] B. Zhou, D. Xu, C. Li, C. Y. Chung, Y. Cao, K. W. Chan, and Q. Wu, "Optimal scheduling of biogas–solar–wind renewable portfolio for multicarrier energy supplies," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6229–6239, Nov. 2018, doi: 10.1109/TPWRS.2018.2833496.

[6] S. Wang, G. Geng, J. Ma, Q. Jiang, H. Huang, and B. Lou, "Operational bottleneck identification based energy storage investment requirement analysis for renewable energy integration," *IEEE Trans. Sustain. Energy*, vol. 12, no. 1, pp. 92–102, Jan. 2021, doi: 10.1109/TSTE.2020.2984651.

[7] C. Luo, Y.-F. Huang, and V. Gupta, "Stochastic dynamic pricing for EV charging stations with renewable integration and energy storage," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1494–1505, Mar. 2018, doi: 10.1109/TSG.2017.2696493.

[8] W. Huang, N. Zhang, J. Yang, Y. Wang, and C. Kang, "Optimal configuration planning of multi-energy systems considering distributed renewable energy," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1452–1464, Mar. 2019, doi: 10.1109/TSG.2017.2767860.

[9] C. Chen, Z. Wu, and Y. Zhang, "The charging characteristics of large-scale electric vehicle group considering characteristics of traffic network," *IEEE Access*, vol. 8, pp. 32542–32550, 2020, doi: 10.1109/ACCESS.2020.2973801.

[10] C. Mu, W. Liu, and W. Xu, "Hierarchically adaptive frequency control for an EV-integrated smart grid with renewable energy," *IEEE Trans. Ind. Informat.*, vol. 14, no. 9, pp. 4254–4263, Sep. 2018, doi: 10.1109/TII.2018.2846545.

[11] L. Gan, X. Chen, K. Yu, J. Zheng, and W. Du, "A probabilistic evaluation method of household EVs dispatching potential considering Users' multiple travel needs," *IEEE Trans. Ind. Appl.*, vol. 56, no. 5, pp. 5858–5867, Sep. 2020, doi: 10.1109/TIA.2020.2989690.

[12] M. Giansoldati, A. Monte, and M. Scorrano, "Barriers to the adoption of electric cars: Evidence from an Italian survey," *Energy Policy*, vol. 146, Nov. 2020, Art. no. 111812, doi: 10.1016/j.enpol.2020.111812.

[13] M. Shafie-khah, E. Heydarian-Forushani, G. J. Osorio, F. A. S. Gil, J. Aghaei, M. Barani, and J. P. S. Catalao, "Optimal behavior of electric vehicle parking lots as demand response aggregation agents," *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2654–2665, Nov. 2016, doi: 10.1109/TSG.2015.2496796.
[14] S. Wolff and R. Madlener, "Driven by change: Commercial drivers' acceptance and efficiency perceptions of light-duty electric vehicle usage in Germany," *Transp. Res. C, Emerg. Technol.*, vol. 105, pp. 262–282, Aug. 2019, doi: 10.1016/j.trc.2019.05.017.
[15] L. Chen, Z. Pan, T. Yv, and K. Wang, "Real-time optimal dispatch for large-scale electric vehicles based on dynamic non-cooperative game theory," *Automat. Electr. Power Syst.*, vol. 43, no. 24, pp. 32–40 and 66, Jul. 2019, doi: 10.7500/AEPS20181116007.
[16] M. S. Islam, N. Mithulananthan, and D. Q. Hung, "Coordinated EV charging for correlated EV and grid loads and PV output using a novel, correlated, probabilistic model," *Int. J. Electr. Power Energy Syst.*, vol. 104, pp. 335–348, Jan. 2019, doi: 10.1016/j.ijepes.2018.07.002.
[17] M. Shin, D.-H. Choi, and J. Kim, "Cooperative management for PV/ESS-enabled electric vehicle charging stations: A multiagent deep reinforcement learning approach," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3493–3503, May 2020, doi: 10.1109/TII.2019.2944183.
[18] Q. Huang, Q.-S. Jia, Z. Qiu, X. Guan, and G. Deconinck, "Matching EV charging load with uncertain wind: A simulation-based policy improvement approach," *IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1425–1433, May 2015, doi: 10.1109/TSG.2014.2385711.
[19] M. A. A. Abdalla, W. Min, and O. A. A. Mohammed, "Two-stage energy management strategy of EV and PV integrated smart home to minimize electricity cost and flatten power load profile," *Energies*, vol. 13, no. 23, p. 6387, Dec. 2020, doi: 10.3390/en13236387.
[20] K. Chaudhari, N. K. Kandasamy, A. Krishnan, A. Ukil, and H. B. Gooi, "Agent-based aggregated behavior modeling for electric vehicle charging load," *IEEE Trans. Ind. Informat.*, vol. 15, no. 2, pp. 856–868, Feb. 2019, doi: 10.1109/TII.2018.2823321.
[21] B.-S. Chen, W.-Y. Chen, C.-T. Yang, and Z. Yan, "Noncooperative game strategy in cyber-financial systems with Wiener and Poisson random fluctuations: LMIs-constrained MOEA approach," *IEEE Trans. Cybern.*, vol. 48, no. 12, pp. 3323–3336, Dec. 2018, doi: 10.1109/TCYB.2018.2869018.
[22] Z. Xiao, X. Shen, F. Zeng, V. Havyarimana, D. Wang, W. Chen, and K. Li, "Spectrum resource sharing in heterogeneous vehicular networks: A noncooperative game-theoretic approach with correlated equilibrium," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9449–9458, Oct. 2018, doi: 10.1109/TVT.2018.2855683.
[23] L. Cheng and T. Yu, "Game-theoretic approaches applied to transactions in the open and ever-growing electricity markets from the perspective of power demand response: An overview," *IEEE Access*, vol. 7, pp. 25727–25762, 2019, doi: 10.1109/ACCESS.2019.2900356.
[24] Y. Huang, N. Peng, Y. Lin, J. Fan, Y. Zhang, and Y. Y, "SAC reinforcement learning based multi-agent dynamic resource allocation in Internet of Vehicles," *Comput. Eng.*, vol. 2020, no. 12, pp. 1–12, Dec. 2020. [Online]. Available: http://kns.cnki.net/kcms/detail/31.1289.TP.20201210.1648.003.html
[25] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, pp. 2976–2989.
[26] F. Fu, Y. Kang, Z. Zhang, F. R. Yu, and T. Wu, "Soft actor–critic DRL for live transcoding and streaming in vehicular fog-computing-enabled IoV," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1308–1321, Feb. 2021, doi: 10.1109/JIOT.2020.3003398.
[27] T.-H. Teng, A.-H. Tan, and J. M. Zurada, "Self-organizing neural networks integrating domain knowledge and reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 5, pp. 889–902, May 2015, doi: 10.1109/TNNLS.2014.2327636.
[28] J. Dong, W. Wu, Y. Gao, X. Wang, and P. Si, "Deep reinforcement learning based worker selection for distributed machine learning enhanced edge intelligence in Internet of vehicles," *Intell. Converged Netw.*, vol. 1, no. 3, pp. 234–242, Dec. 2020, doi: 10.23919/ICN.2020.0015.
[29] N. Bouhlel and A. Dziri, "Kullback–Leibler divergence between multivariate generalized Gaussian distributions," *IEEE Signal Process. Lett.*, vol. 26, no. 7, pp. 1021–1025, Jul. 2019, doi: 10.1109/LSP.2019.2915000.

DUNNAN LIU received the B.E. and Ph.D. degrees in electrical engineering from Tsinghua University, China. 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.

LINGXIANG WANG received the B.S. degree in business administration from North China Electric Power University, China, in 2019, where she is currently pursuing the master's degree. Her research interests include power market analysis and power load management.

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

HUA LI received the bachelor's degree, in 2019. He is currently pursuing the master's degree with the School of Economics and Management, North China Electric Power University. His main research interest includes electricity market.

MINGGUANG LIU received the master's degree from North China Electric Power University, in 2019, where he is currently pursuing the Ph.D. degree with the School of Economics and Management. His main research interests include low-carbon and energy economy development and comprehensive energy systems.

XIAOFENG XU received the master's degree from North China Electric Power University, in 2020, where he is currently pursuing the Ph.D. degree with the School of Economics and Management. His main research interests include low-carbon and energy economy development and comprehensive energy systems.