IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 12, 2021, accepted January 15, 2021, date of publication January 29, 2021, date of current version February 8, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3055517

Dynamic Pricing Strategy of Electric Vehicle Aggregators Based on DDPG Reinforcement Learning Algorithm

DUNNAN LIU, WEIYE WANG, LINGXIANG WANG, HEPING JIA, AND MENGSHU SHI
School of Economics and Management, North China Electric Power University, Beijing 102206, China

Corresponding author: Heping Jia (jiaheping@ncepu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 72001078, and in part by the Power System State Key Laboratory under Grant SKLD20M12.

ABSTRACT The fixed service charge pricing model adopted by traditional electric vehicle aggregators (EVAs) is difficult to effectively guide the demand side resources to respond to the power market price signal. At the same time, real-time pricing strategy can flexibly reflect the situation of market supply and demand, shift the charging load of electric vehicles (EVs), reduce the negative impact of disorderly charging on the stable operation of power systems, and fully tap the economic potential of EVA participating in the power market. Based on the historical behavior data of EVs, this paper considers various market factors such as peak-valley time-of-use tariff, demand-side response mode and deviation balance of spot market to formulate the objective function of EVA comprehensive revenue maximization and establishes a quarter-hourly vehicle-to-grid (V2G) dynamic time-sharing pricing model based on deep deterministic policy gradient (DDPG) reinforcement learning algorithm. The EVA yield difference between peak-valley time-of-use tariff and hourly pricing strategy under the same algorithm is compared through the case studies. The results show that the scheme with higher pricing frequency can guide the charging behavior of users more effectively, tap the economic potential of power market to a greater extent, and calm the load fluctuation of power grid.

INDEX TERMS Electric vehicle aggregator, dynamic pricing, DDPG algorithm, charging behavior guidance.

I. INTRODUCTION

With the development of electric vehicles (EVs), electric vehicles have gradually become new transaction subject in power market with the continuous expansion of its volume. With the dual characteristics of power supply and load, electric vehicles can increase or decrease the charging power or transfer the charging time according to the needs of the power grid, or even send power back to the power grid through V2G technology [1]–[3]. Compared with the traditional unidirectional flexible load which is passively controlled, EVs have more active control ability [4], [5], thus significantly improving the cost efficiency of the power system [6], [7].

Since individual EVs do not have the ability to participate independently in the power wholesale market, to realize their flexibility potential, fragmented EVs need to be effectively aggregated as a whole to participate in the power market. The most promising mechanism to achieve this kind of market integration is EV aggregation (EVA) [8]. EVA represents a large number of EVs in the retail market and coordinates its behavior characteristics according to the market situation and the operation of EVs to maximize its overall profit.

Since it is difficult for demand-side resources to be directly controlled by scheduling, the price-adjusted coordination approach has recently attracted a lot of attention [9]. EVA provides fixed prices for different periods through the analysis of market information, and users independently determine their optimal charge-discharge response through their own value function, so as to solve the problem of their electricity efficiency maximization.

At present, EVA in China mainly benefits from the charging price composed of grid benchmark electricity price and charging service fee. The fixed charging service fee has been unable to cope with the flexible power market environment. In order to fully activate the flexibility of EVs and tap the potential of the power market, the pricing strategy of

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

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

aggregators needs to be dynamic pricing in periods. Numerical tests show that a real-time smart-charging method (N-RT) that both considers currently connected EVs and future prediction of the EVs plugged in highly improves valley-filling and economic benefit under various levels of prediction accuracy [10].

In order to achieve real-time pricing of EVA, dynamic planning is the most commonly used decision-making method [11], [12]. However, when faced with high-dimensional and high-frequency refined pricing problems, dynamic programming method requires enormous computing resources. At the same time, dynamic programming methods cannot cope with the impact of unstable environments brought by users with highly flexible behaviors.

In this case, EVA needs to design an appropriate mechanism to design the pricing scheme to maximize its overall profit while considering the user response mode.

In terms of EVA pricing strategy research, existing pricing strategies can be classified into four categories: 1) cost analysis [13]–[15], which takes the purchasing cost of EVA power plus reasonable profit as the declared electricity price; 2) user behavior analysis [16]–[18], where EVA determines the pricing according to the predicted user elasticity coefficient of the price according to EV travel rules; 3) game theory [19], [20], which mainly includes three steps, that is, constructing a game model according to charging transactions, searching for the equilibrium point of the model, and determining the optimal bidding strategy of power generators; 4) intelligent optimization algorithms, such as competitive co-evolution algorithm [21], [22], fuzzy self-adaptive search algorithm [23] and reinforcement learning (RL) method [24], [25], etc.

Among many pricing strategies, generation cost analysis is the basis of EVA pricing. This method is simple and feasible, but it does not take into account the price information of the external environment, and it is difficult to maximize its own profit [7]; after solving the constraints, user behavior analysis usually gets the theoretically fixed user elasticity coefficient or probability density curve, which cannot accurately fit the daily dynamic travel demands of users. The game theory method has a significant advantage in solving perfect information game problem under fixed environment, but its effect on imperfect information processing under variable environment is not ideal [19], [20]. Due to its own limitations, the traditional intelligent optimization algorithm has high computational complexity to solve the optimization problem under the real-time decision.

Reinforcement learning method is a data-driven intelligent control method, which can provide the optimal control strategy in real time according to the changes of system environment through the mechanism of ''action to reward''. its model-free characteristic can avoid the complex system modeling process under certain conditions [26]. Traditional reinforcement learning method based on value function selects the action with the highest return as the strategy by comparing the size of different environment-action pairs, so it is necessary to discretize the control action. However, the degree and method of discretization may have a great influence on the final result, and the complexity of action space will increase geometrically with the increase of actions, which limits its use in complex decision problems. Deep deterministic policy gradient(DDPG) algorithm based on actor-critic system is another kind of reinforcement learning algorithm, in which the agent learns the mapping function between environment and action, and does not need to discretize the control action [25], so it has great application potential in solving complex problems.

Therefore, aiming at the pricing problem of tracking 15-minute multi-market price information, this paper analyzes the user's historical behavior data and uses DDPG algorithm to build a quarter-hourly dynamic pricing model in the real-time continuous strategy space, which overcomes the defects of the traditional algorithm. Through the comparative analysis of different pricing frequencies, the effectiveness of the algorithm and the necessity of fine pricing are verified.

II. ANALYSIS OF MARKET TRADING PATTERNS OF ELECTRIC VEHICLES

With the continuous development of China's power market, the variety of electricity trading is gradually enriched, and demand-side response has gradually become an important supplement to the stability of the total load of the system. The variety of power trading leads to the difference and volatility of power use value in different power periods.

Under the peak-valley time-of-use tariff, the average peak-valley time-of-use tariff can reach about 3 times. Demand side response transactions conducted at specific times can bring additional benefits to EVAs. Meanwhile, the development of China's spot market in the future will eliminate the traditional deviation assessment method, and the uncertainty of EV users' electricity consumption will make up for the forecast deviation value with the dynamic price of the spot market. The translatability of EVs has great economic potential and also faces great challenges in the future power market.

The research of this paper includes three stakeholders: power grid company, EVA and EV. The interaction framework is illustrated in the Figure 1. The grid company guarantees the power supply of EVA, provides price information and scheduling demand at different periods, and encourages EVA to optimize EVs power purchase behavior to reduce peak and valley difference. EVA provides corresponding services for the demand side of the power grid, such as peak shaving and valley filling, and makes up for the extra power difference in the spot market, which is brought by the demand-side response.

At the same time, EV inputs the market price information and users' travel patterns into the DDPG algorithm, outputs EV charging standards, and trades with power grid companies and EV users. EVA's revenue includes three parts. First, EVA time-sharing purchases electricity from the grid company, and earns profits through charging service fee markup. Second, EVA reduces the peak and valley difference of the

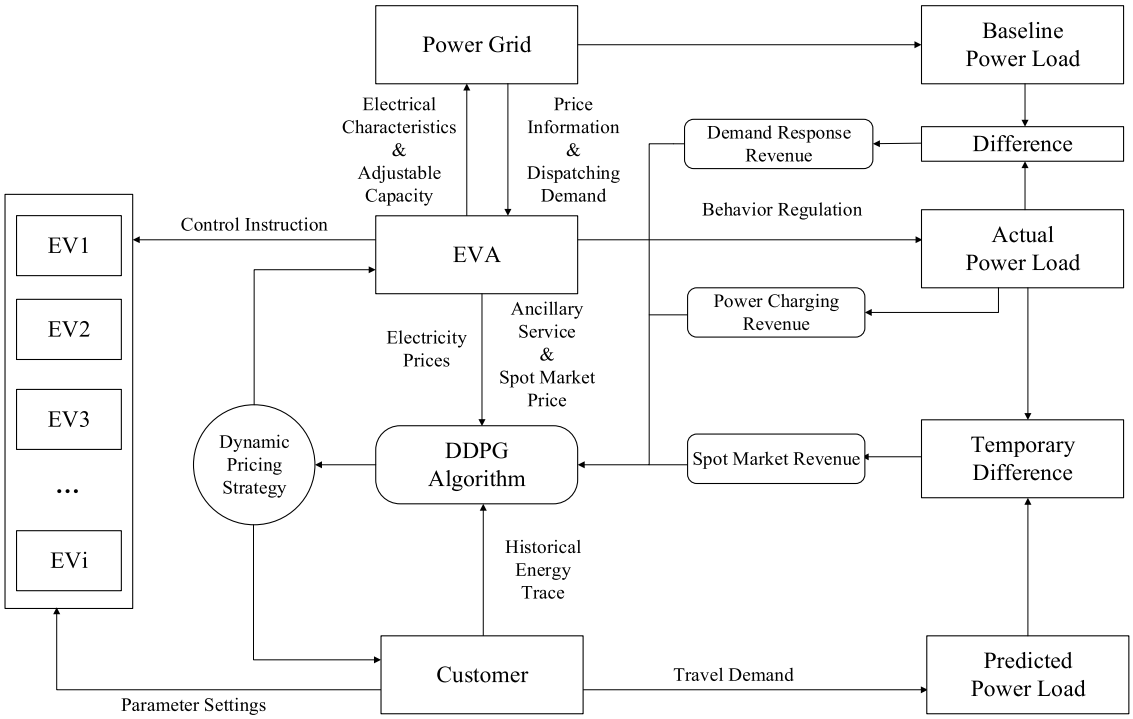

FIGURE 1. Interaction framework of power grid company, EVA and EV.

power grid by realizing orderly charging of EVs, so as to obtain the benefits of the demand side response of the power grid. Finally, EVA eliminates the electric quantity deviation of its response in the spot market, so as to obtain the transaction income of the shifted deviated electric quantity.

In order to develop a daily real-time dynamic pricing strategy, in this paper, EVA builds an iterative DDPG algorithm model with the objective of participating in the demand side response market, the comprehensive revenue maximization of spot market and electricity tariff revenue.

(1) Demand-side response market: The demand-side response market conducts valley filling trade at 26 time points from 0:45 to 7:00 every day; Peak trading starts at 14 hours between 12:45 and 16:00. EVA takes the difference between the reference load curve and the actual load curve confirmed by its own predicted electric energy characteristics and adjustable capacity as the settlement basis, and the calculation formula is as follows:

$$R^a = \sum_{i=3}^{28} \left(Q_i^{Acutal} - Q_i^{Baseline}\right) p_i^a t + \sum_{i=51}^{64} \left(Q_i^{Baseline} - Q_i^{Acutal}\right) p_i^a t \quad (1)$$

where, R^a is the demand side response market return, Q_i^{Acutal} is the real load value at a certain point, $Q_i^{Baseline}$ is the benchmark load value at a certain point, p_i^a is the demand side response clearing price at a certain point, and t is charging duration of each time point.

(2) Spot market: The deviation value between EVA's estimated load curve and the actual curve at each time point according to user travel demand and contract information needs to be eliminated in the spot market. The spot market earnings are calculated as follows:

$$R^s = \sum_{i=1}^{96} \left(Q_i^{Predict} - Q_i^{Actual}\right) p_i^s t \quad (2)$$

where, R^s is the spot market yield, $Q_i^{Predict}$ is the expected load value at a certain point, and $p_i^s$is the clearing price in the spot market at a certain point.

(3) Electricity fee income: EVA electricity fee income is divided into two parts: time-sharing power purchase cost and charging service fee. The time period of time-of-use electricity price is divided into: the peak time period is 12:00-21:00; The normal period is 7:00-12:00 and 21:00-24:00; The valley period is 0:00-7:00. The calculation formula is as follows:

$$R^c = \sum_{i=1}^{28} Q_i^{Actual} \left(p_i^c + p^L\right) t + \left(\sum_{i=28}^{48} Q_i^{Actual} + \sum_{i=84}^{96} Q_i^{Actual}\right) \left(p_i^c + p^M\right) t + \sum_{i=48}^{84} Q_i^{Actual} \left(p_i^c + p^H\right) t \quad (3)$$

where, R_c is the electricity fee income, p_i^c is the dynamic service fee at a certain point, p_i^H is the electricity purchase

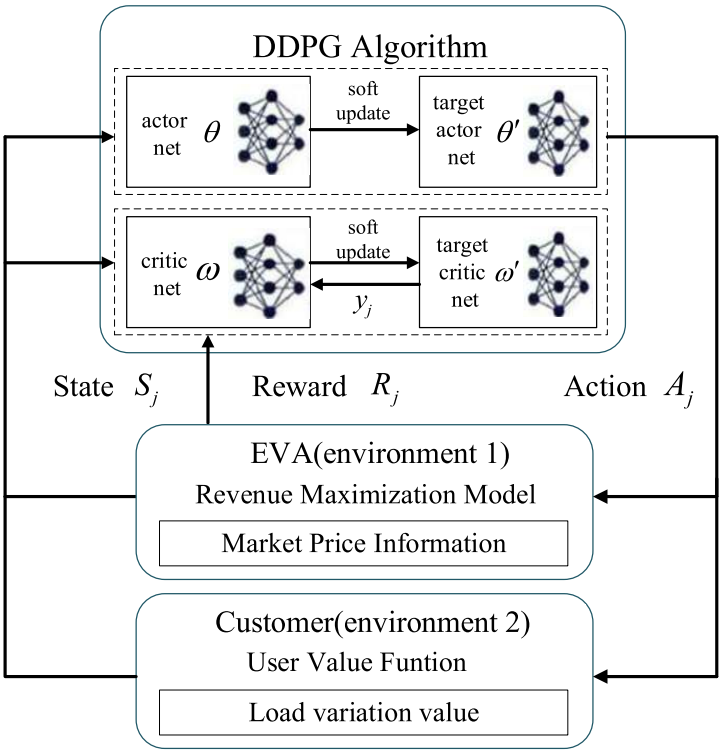

FIGURE 2. Calculation step of DDPG.

price during the peak period, p_i^M is the electricity purchase price during the normal period, and p_i^L is the electricity purchase price during the valley period.

Since EV users have different travel patterns, EVA can adjust the electricity price and other incentive measures to enable EV users to adjust their electricity consumption behavior autonomically, which bases on the user's demand elasticity, and indirectly complete the optimization of power grid operation and its own benefits.

The loss function of reality network is as follows [27]:

$$J(\omega) = \frac{1}{m}\sum_{j=1}^{m}\left(y_j - Q\left(\phi\left(S_j\right), A_j, \omega\right)\right)^2 \tag{4}$$

In each iteration, m samples are sampled from the experience playback set D. Calculate the current target Q value:

$$y_j = \begin{cases} R_j & is\_end_j\ is\ true \\ R_j + \gamma Q\left(\phi\left(S_{j+1}\right), A_{j+1}, \omega\right) & is\_end_j\ is\ false \end{cases} \tag{5}$$

where, m is the number of samples, $Q\left(\phi\left(S_j\right), A_j, \omega\right)$ is the action A_j taken by the sample in the state, the action value is calculated by the reality Critic network. And y_j is the target action value calculated by the target Critic. R_j is the reward in the state of S_j taking action A_j, and γ is the discount factor.

III. PRINCIPLE OF DDPG REINFORCEMENT LEARNING ALGORITHM

reinforcement learning uses the reward and punishment information obtained from the interaction between the agent and the environment to guide the agent's behavior. DDPG is an actor-critic based algorithm in reinforcement learning [28]. Its core idea is to use Actor network to generate the behavioral strategy of agents. The Critic network can judge the quality of actions and guide the updating direction of actions [26].

As shown in Figure 2, the DDPG structure contains an Actor network with parameter θ and an Critic network with parameter ω, which are brave to calculate deterministic strategy $\pi(s|\theta)$ and action value function $Q(s, a|\omega)$ respectively. Because the single network learning process is unstable, it draws on the successful experience of DQN fixed target network and subdivides Actor network and Critic network into one reality network and one target network respectively. The real network and the target network structure is the same, the target network parameters to determine the frequency by the real network parameters soft update.

The visual environment of each step of the DDPG algorithm is the market price information obtained from EVA at each time point, and the load change information obtained from users. After feeding the environment information into Actor Net, the model outputs the dynamically priced action value of quarter-hourly.

EVA, based on price information in different markets, uses the difference between the output dynamic pricing and fixed

pricing income as the reward value, with which and environmental information being imported into the critic network. EVA will be used by the critic network as a measure of how good the pricing strategy will be. The bigger the revenue, the better the strategy, allowing actor networks to adjust network weights based on the feedback from the critic network.

According to the output price change information and its own value function, the user outputs the load change value under the new price information, which, together with the new round of market price information received by EVA, becomes the environmental input value of the new step. Execute the steps until the end of the mock business round.

When the training reaches a certain number of rounds, the parameters of Target Net will be updated on a sliding average based on the actual value of the real network parameters.

The loss function of reality network is as follows [27]:

$$J(\omega) = \frac{1}{m}\sum_{j=1}^{m}\left(y_j - Q\left(\phi\left(S_j\right), A_j, \omega\right)\right)^2 \tag{6}$$

In each iteration, m samples are sampled from the experience playback set D. The current target value is calculated as follows:

$$y_j = \begin{cases} R_j & is\_end_j\ is\ true \\ R_j + \gamma Q\left(\phi\left(S_{j+1}\right), A_{j+1}, \omega\right) & is\_end_j\ is\ false \end{cases} \tag{7}$$

where, m is the number of samples, $Q\left(\phi\left(S_j\right), A_j, \omega\right)$ is the action taken by the sample in the state S_j, the action value calculated by the reality Critic network; and y_j is the target action value calculated by the target Critic. R_j is the reward obtained by the sample taking an action A_j in the state of S_j, and γ is the discount factor.

The loss gradient for a real Actor Net is as follows:

$$\nabla J(\theta) = -\frac{1}{m}\sum_{j=1}^{m}\left(\nabla_a Q\left(s_i, a_i, \omega\right)\right) \tag{8}$$

where, the gradient descent method is used to find the minimum value of the loss function $J(\theta)$, which is equivalent to the process of maximizing the action value $Q\left(\phi\left(S_j\right), A_j, \omega\right)$.

At the end of each iteration, the Critic Target Net and Actor Target Net parameters are updated in the following manner:

$$\omega' \leftarrow \tau\omega + (1-\tau)\,\omega' \tag{9}$$

$$\theta' \leftarrow \tau\theta + (1-\tau)\,\theta' \tag{10}$$

If S_j is the termination state, the round is iterated out.

IV. DYNAMIC TIME-SHARING PRICING MODEL FOR EV CHARGING

A. CHARGING PRICE ELASTICITY COEFFICIENT OF USERS BASED ON VALUE FUNCTION

Under the dynamic adjustment of multi-period electricity price, the charging behavior of users in the current period will be comprehensively affected by the electricity price in different time periods.

By summarizing charging information for all individual user in different battery types, this paper focus on the relationship between the change of charging price and overall charging load in different time period, then studies the relationship based on value function. We use users load data to different price record from dynamic battery models to estimate unknown parameters, and then calculate the elasticity coefficient.

According to the Generalized Leontief value function [29], the total cost of an electric car can be expressed in the following equation with the charging price and charging demand in different time periods:

$$C = g\left(q_i\right)\sum_{i=1}^{96}\sum_{j=1}^{96}\omega_{ij}\sqrt{\left(p_i p_j\right)} \tag{11}$$

where, i, j represents different time periods, C represents the total charging cost of EV, $g\left(q_i\right)$ represents charging demand of EV, p_i and p_j respectively represent charging price of EV in time period i and time period j.

The ratio σ_i of fixed period expenses to total expenses can be expressed as:

$$\sigma_i = \frac{\sum_j \omega_{ij}\sqrt{\left(p_i p_j\right)}}{\sum_{k=1}^{96}\sum_j \omega_{kj}\sqrt{\left(p_k p_j\right)}} \tag{12}$$

Thus, the self-elastic coefficient ε_{ii} and the mutual elastic coefficient ε_{ij} of the user's charging behavior can be calculated as follows:

$$\varepsilon_{ii} = \frac{1}{2}\left(\frac{\omega_{ii}}{\sum_{ki}\omega_{ik}}\sqrt{\frac{p_i}{p_k}} - 1\right) \tag{13}$$

$$\varepsilon_{ij} = \frac{1}{2}\left(\frac{\omega_{ij}}{\sum_{ki}\omega_{ik}}\sqrt{\frac{p_j}{p_k}}\right) \tag{14}$$

B. DYNAMIC TIME-SHARING PRICING MODEL BASED ON DDPG REINFORCEMENT LEARNING

The purpose of electricity price guidance is to improve the potential economic benefits of EV loaders, reduce the peak and valley difference of the power grid, and calm the load fluctuation of the power grid. The dynamic pricing model of EV charging can be established by using the comprehensive revenue of demand side response market, spot market and charging price to reflect the income change, and at the same time considering the economic efficiency and load change of EV users. Select the maximum comprehensive income of the calculation period as the objective function:

$$\max R = \sum_i \left(R^a + R^s + R^c\right) \quad i = 1, 2, \ldots, n \tag{15}$$

In order to ensure the overall rationality of dynamic pricing strategy and avoid the damage of excessive profit to users'

interests under greedy algorithm, the pricing model should meet the following constraints:

(1) The earnings of a single trading day under the dynamic pricing strategy should not be lower than the earnings under the fixed pricing strategy:

$$R^{a'} = \sum_{i=3}^{28} \left(Q_i^O - Q_i^{Baseline}\right) p_i^a t + \sum_{i=51}^{64} \left(Q_i^O - Q_i^{Baseline}\right) p_i^a t \quad (16)$$

$$R^{s'} = \sum_{i=1}^{96} \left(Q_i^{Predict} - Q_i^O\right) p_i^s t \quad (17)$$

$$R^{c'} = \sum_{i=1}^{28} Q_i^O \left(p_0 + p^L\right) t + \left(\sum_{i=28}^{48} Q_i^O + \sum_{i=84}^{96} Q_i^O\right) \left(p_0 + p^M\right) t + \sum_{i=48}^{84} Q_i^O \left(p_0 + p^H\right) t \quad (18)$$

$$R_i^a + R_i^s + R_i^c \geq R_i^{a'} + R_i^{s'} + R_i^{c'} \quad i = 1, 2, \ldots, n \quad (19)$$

wherein, Q_i^O refers to the load in the period before dynamic electricity price is implemented, $R^{a'}$ refers to the demand side response market income before dynamic electricity price is implemented, $R^{s'}$ refers to the spot market income before dynamic electricity price is implemented, $R^{c'}$ refers to the charging income before dynamic electricity price is implemented, and p_0 refers to the fixed service fee before dynamic electricity price is implemented.

(2) After the implementation of dynamic electricity price, the total electricity load of EV users should remain unchanged:

$$\sum_j \sum_i^{96} Q_i^O = \sum_j \sum_i^{96} Q_i^{Actual} \quad j = 1, 2, \ldots, n \quad (20)$$

(3) After the implementation of dynamic electricity price, the user's total charging cost should not be higher than the charging cost under the fixed service fee:

$$\sum_j \sum_i^{96} Q_i^O p_i^c t \leq \sum_j \sum_i^{96} Q_i^{Actual} p_0 t \quad j = 1, 2, \ldots, n \quad (21)$$

(4) Considering the economic limitations on the user side and the grid side, the charging service fee should be changed within a certain range:

$$p_i \in (p_{\min}, p_{\max}) \quad (22)$$

V. SOLUTION OF DYNAMIC TIME-SHARING PRICING MODEL BASED ON DDPG REINFORCEMENT LEARNING

Input: Actor Current Net, Actor Target Net, Critic Current Net, Critic Target Net, parameters are θ, θ', ω, ω', attenuation factor γ, soft update coefficient τ, sample number of batch gradient descent m, target Q network parameter update frequency C, maximum number of iterations T, random noise function N.

Output: the best Actor Current Net parameters θ, Critic Current network parameters ω.

Implement the framework of DDPG, as shown in the figure below:

1. Random initialization $\theta, \omega, \omega' = \omega, \theta' = \theta$. Clear the set of experience replays D,

2. Iterate from the first step of the first round,

A) Initialize as the first state S of the current state sequence and get its eigenvector $\phi(S)$.

B) The current network in Actor $A = \pi_\theta(\phi(S)) + N$ gets an action based on the state$S$,

C) Perform actions A, get new status S', reward R, terminate status or not,

D) Store the quintuple $\{\phi(S), A, R, \phi(S'), is\_end\}$ into the experience playback set D,

E) $S = S'$,

F) Sample m samples from the experience playback set D and calculate the current target Q value y_j:

$$y_j = \begin{cases} R_j & is\_end_j\ is\ true \\ R_j + \gamma Q\left(\phi\left(S_{j+1}\right), A_{j+1}, \omega\right) & is\_end_j\ is\ false \end{cases} \quad (23)$$

G) Use the mean square error loss function $\frac{1}{m}\sum_{j=1}^{m}\left(y_j - Q\left(\phi\left(S_j\right), A_j, \omega\right)\right)^2$, to update all parameters of the current network through gradient back propagation of the neural network,

H) Use $J(\theta) = -\frac{1}{m}\sum_{j=1}^{m} Q(s_i, a_i, \theta)$ to update all parameters of the Actor's current network through gradient back propagation of the neural network.

I) If $T\%C = 1$, then update the target network and Actor target network parameters:

$$\omega' \leftarrow \tau\omega + (1-\tau)\omega' \quad (24)$$

$$\theta' \leftarrow \tau\theta + (1-\tau)\theta' \quad (25)$$

J) If S_{j+1} is the termination state, the current round iteration is completed, otherwise go to step B) [27].

The relevant parameters of the model are as table 1.

VI. CASE ANALYSIS

Based on the training data of 166-day actual charging of EV in a certain region of northern Hebei Province, this paper studies EVA profit growth under different pricing strategies. The charging power is calculated according to 3 kW, and the user's demand response will shift the whole charging behavior of the user. When the time-of-use electricity price is implemented, the fixed charging service fee of EV in this region is 0.5 yuan/(kW·h), and the pricing range of EV service fee is set from 0 yuan/(kW·h) to 1 yuan/(kW·h). The period price of charging electricity purchased from the grid is

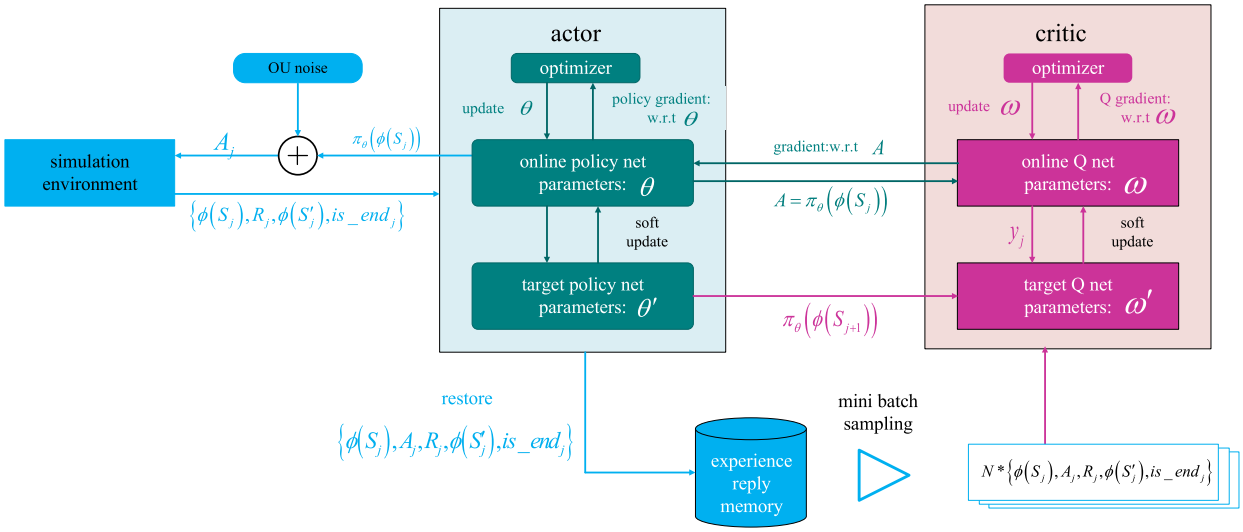

FIGURE 3. Implementation framework for DDPG.

TABLE 1. Main parameters table.

parameters	symbols	numerical
learning rate for actor	α_a	0.001
learning rate for critic	α_b	0.002
reward discount	γ	0.9
soft replacement	tau	0.01
Memory capacity	D	10000
update batch size	m	32
max episodes	P	2500
Soft replacement rate	C	100
max nums of iterations	T	365
epsilon greedy value	ε	0.01

divided into three parts. the peak period is 12:00-21:00; The normal period is 7: 00-12: 00 and 21: 00-24:00 and the valley time is 0: 00-7: 00. The electricity price in the peak period is 1.29 yuan/(kW·h), the electricity price in the normal period is 0.87 yuan/(kW·h), and the electricity price in the valley period is 0.46 yuan/(kW·h). The trading frequency of spot market and demand-side responsive market is 15 minutes, and the developing time of demand-side responsive market is 0:45-7:00 and 12:45-21:00.

In order to verify the effectiveness of the proposed method, the DDPG algorithm is used to access the clean price for demand side response market and spot market during the corresponding load data period for simulation. In the simulation environment, taking 166 days as a training cycle, the daily EV service fees within the training cycle are respectively priced by peak-valley time-of-use tariff, hour pricing and quarter-hourly pricing. the total revenue and load changes of EVA after each training cycle are calculated for comparison.

The reinforcement learning model in the experiment is implemented by Python Tensorflow 2.0. Constraints in formulas (13) - (20) are calculated using the Python interface in Xpress Optimizer. The experimental computer model is a quad-core 2.60-ghz Intel Core I7-6700HQ processor with 16GB of memory.

The study compared the algorithm convergence of three scenarios and the overall EVA revenue change in the case of daily updating of the pricing strategy. The three scenarios were respectively the overall operation situation of 166 days under the quarter-hourly pricing strategy, peak-valley time-of-use tariff and hourly pricing strategy. The convergence of the algorithm and the revenue of the scenario are shown in Figure 4.

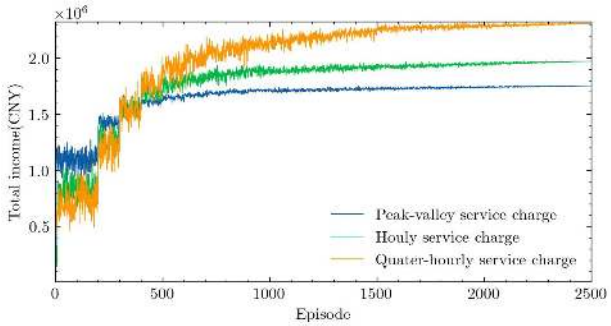

FIGURE 4. Convergence of the algorithm and revenue of the scenario.

Under the setting of daily pricing update, the iterative decision space of the algorithm is large, and the algorithm has some oscillation in the process of convergence. However, after adding the constraint that the return under the update strategy should not be lower than the original return, the algorithm avoided the disorderly iteration starting from negative return. All the three strategies improved the overall return of EVA from the initial round and rose steadily to be stable.

Due to the low output dimension of the algorithm, peak-valley time-of-use tariff can obtain higher extra income from

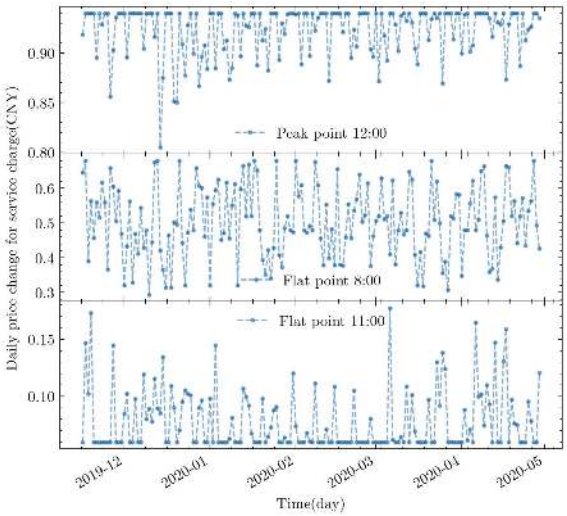

FIGURE 5. The distribution of the average daily service charge under the quarter-hourly pricing of EVA.

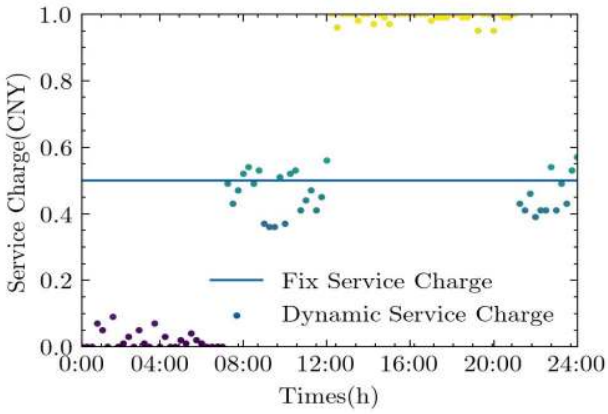

FIGURE 6. Daily price fluctuation at typical time points of quarter-hourly.

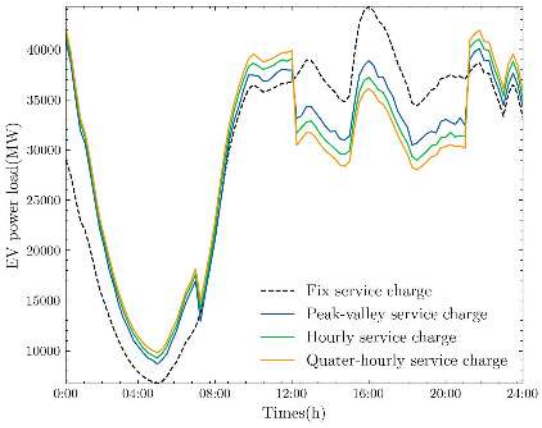

FIGURE 7. Load changes under dynamic pricing, such as EV charging load.

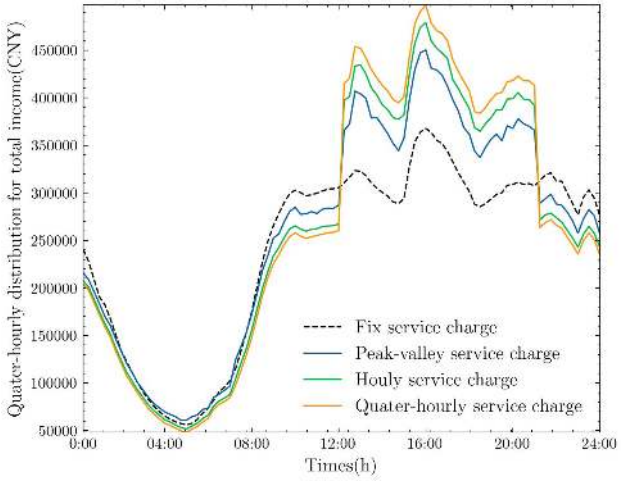

FIGURE 8. Trend change of average return at time point.

the initial round and quickly converge after 500 iterations. However, because the three-stage pricing strategy is relatively simple, there is insufficient room for rise in subsequent iterations. The output of the quarter-hourly pricing strategy is relatively complex, and the additional revenue obtained in the initial iteration is low, and it tends to converge after 1500 iterations. However, this strategy achieves the highest additional revenue of the three pricing strategies (¥2.31 million RMB), which is 1.18 times of the hourly pricing strategy and 1.31 times of the peak-valley time-of-use tariff.

Figure 5 shows the distribution of the average daily service charge under the quarter-hourly pricing of EVA. The price distribution and mean values of the three pricing strategies are approximately similar under the constraint of no increase in total electricity consumption. In the valley period and peak period of demand side response trading, the pricing results under the DDPG algorithm are all close to the extreme of the lower and upper limit of the price. However, when there is no demand side response transaction in normal period, the average price of the price in the valley period under the quarter-hourly pricing method is slightly lower, and the volatility is relatively high.

Meanwhile, the daily price fluctuation at typical time points of quarter-hourly is shown in Figure 6. It can be found in Figure 6 that the trend of price volatility at the time point is the same as that of the corresponding period. In peak period and valley period, the price fluctuation range is lower and the fluctuation trend is slower, while in ordinary period, the price fluctuation range is larger and the fluctuation trend is steeper.

Table 2 shows the mean and standard deviation of quarter-hourly pricing and other pricing strategies at different time periods. In contrast, the higher the frequency of pricing, the greater the price volatility. The volatility of the normal period is significantly higher than that of other periods, among which the volatility of the normal period under the quarter-hourly pricing is about 3 times that of the rest periods. At the same time, it can be found that under different pricing frequencies, the average price of each period is roughly similar, and the average price of the ordinary period increases slightly with the increase of pricing frequency. Therefore, it

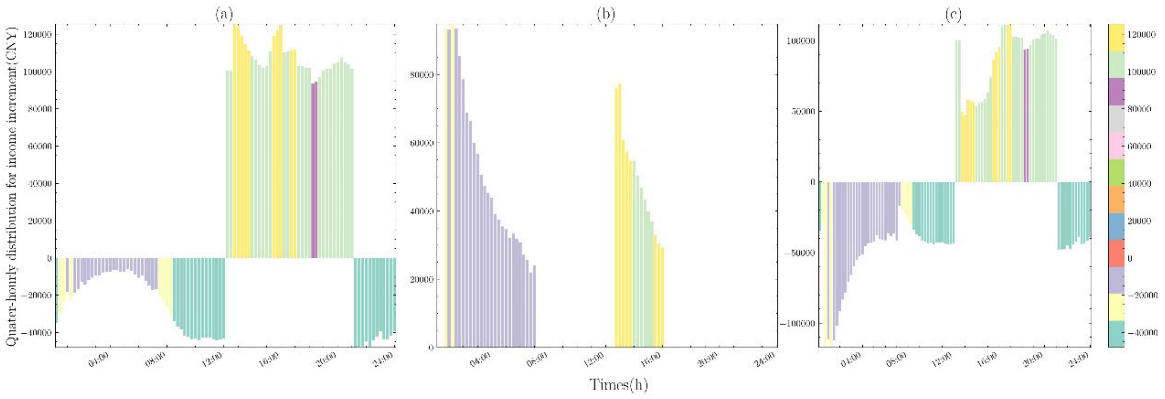

FIGURE 9. Sources of incremental revenue.

TABLE 2. Mean and standard deviation of the three pricing strategies in different time periods.

Pricing strategy		Peak-time	Flat-time	Valley-time
Peak-valley	u	0.976	0.498	0.036
	σ	0	0	0
Hourly	u	0.973	0.475	0.032
	σ	0.016	0.029	0.012
Quarter-hourly	u	0.982	0.461	0.022
	σ	0.024	0.069	0.026

TABLE 3. Load variation range under dynamic pricing, such as EV charging load.

Pricing strategy	Load change (MW)		
	Peak-time	Flat-time	Valley-time
Peak-valley	-172581.43	33596.73	138984.70
Hourly	-215444.87	61882.66	153562.21
Quarter-hourly	-254062.48	87091.26	166971.22

can be concluded that the difference of pricing strategies in ordinary period is the key to the overall income gap.

The advantage of the quarter-hourly pricing strategy is that it can track the price signals of the power market under load adjustment demand over a more continuous action space. In order to verify the physical characteristics of the pricing method under DDPG algorithm, load changes under dynamic pricing are shown in Figure7 and Table 3 of EV charging load. In the case that the total amount of charging load remains unchanged, the load curves of EVs all change greatly in the range of demand side response trading. Among them, the peak load reduction under the quarter-hourly pricing strategy is significantly higher than other pricing strategies, and the overall charging load is reduced by 18% in the peak load of the grid under the quarter-hourly pricing strategy. Among the reduced loads, 34.28% of the loads were divided into normal periods, while the corresponding hourly pricing was 28.72% and the peak-valley time-of-use tariff was 24.17%. The normal section price of the quarter-hourly pricing strategy has high flexibility, which makes the normal section better undertake and transfer the charging load in the peak period and makes the load curve of EVA connected to the power grid more friendly.

After the charging load changes at each time point, the trend change of the average return at the corresponding time point is shown in Figure 8. After the offset of demand side response subsidy and spot market differential power purchase, the income of peak-valley time-of-use tariff was higher than that of fixed pricing, while the income of other pricing methods was slightly lower than that of fixed pricing methods. The revenue of the normal period is lower than the fixed price, which is due to the transfer load of the peak period and affected by the spot market. In these two periods, the yield of the quarter-hourly pricing strategy was at the lowest. In the peak period, the profit value of the three pricing methods is much higher than that of the fixed pricing. The quarter-hourly pricing strategy achieves the highest return in the peak period and realizes the reversal of the comprehensive income.

Since the total load remains unchanged and the user energy cost does not increase, the main sources of incremental revenue are the spot market and the demand side response market, as shown in Figure 9 and table 4 and Figure 9 (a) shows the incremental value of the average EVA comprehensive income at each time point, Figure 9 (b) shows the incremental value of the average demand side response at each time point, and Figure 9 (c) shows the incremental value of the average spot market at each time point. Spot market earnings are derived from the difference in spot market prices at the point in time after load shifts. The amount of cash fluid is large, which is about 2-3 times of the demand side response market. However, since the total load of EV remains unchanged, the net income realized accounts for less than 10% of the total income. Incremental revenue is mainly contributed by

TABLE 4. Sources of incremental revenue.

Pricing strategy		Reward increment(CNY) Peak-time	Flat-time	Valley-time	Total
Peak-valley	R^s	2211088.5	-653214.9	-1450829.7	107043.8
	R^a	517139.1	0	1102659.1	1619798.
Hourly	R^s	2854741.6	-1047224.8	-1645774.7	161742.0
	R^a	596699.9	0	1241207.2	1837907.2
Quarter-hourly	R^s	3368595.1	-1350961.8	-1826777.7	190855.6
	R^a	733650.1	0	1435080.4	2168730.5

the demand side response market. The yield in the peak period was lower than that in the valley period, and the ratio of the yield was about 1:2, and the yield gradually increased with the increase of the pricing frequency. Finally, under the optimal quarter-hourly pricing strategy, EVA's overall operating income increased by 10%.

VII. CONCLUSION

In order to make full use of the flexibility and regulation ability of EV, this paper takes EVA's participation in the transaction revenue maximization of the electric power market as the goal, and solves the dynamic pricing strategy of EVA in the demand response market and spot market. Aiming at the discrete problem of traditional time-sharing pricing model for EV, a quarter-hourly dynamic pricing strategy based on DDPG reinforcement learning algorithm is proposed to fully develop EV scheduling potential. Finally, taking the annual actual travel data of EV in a certain region of North China and the price data of the electric power trading market as an example, the three scenarios of EVA revenue and load changes under quarter-hourly pricing, hourly electricity price and dynamic peak-valley time-of-use tariff are compared, which verifies the superiority of this method.

The results show that the higher the daily pricing frequency is, the more beneficial it is to guide the charging behavior of EV users with high efficiency. Among them, the price of ordinary period is an important factor that leads to the difference of pricing results of different frequencies. The price of normal period is an important factor that leads to the difference of pricing results of different frequencies. In the case that the pricing in both the valley period and the peak period tends to the limit, under the quarter-hourly pricing strategy, the average price in the normal period has greater fluctuations, which more effectively responds to the price signal in the demand-side market by shifting 18.5% of the peak load to the average period and the peak period, and the demand side response market revenue reaches 82.2% of the incremental revenue. At the same time, the flexible characteristics of EV also make EVA earn extra income from the price difference of different periods in the spot market in the process of load translation, and finally, under the constraint of ensuring that the user's energy cost does not increase, the overall operation income of EVA is increased by 10%. Under the quarter-hourly pricing strategy, the revenue per hour pricing strategy increased by 18%, the peak-valley time-of-use tariff increased by 31%, and the peak-load transfer value increased by 19%, and the peak-valley pricing strategy increased by 47%, which more effectively improved the comprehensive income of EVA and translated the overall fluctuation of power grid load.

REFERENCES

[1] C. Liu, K. T. Chau, D. Wu, and S. Gao, "Opportunities and challenges of vehicle-to-home, vehicle-to-vehicle, and vehicle-to-grid technologies," *Proc. IEEE*, vol. 101, no. 11, pp. 2409–2427, Nov. 2013, doi: 10.1109/JPROC.2013.2271951.

[2] B. K. Sovacool, L. Noel, J. Axsen, and W. Kempton, "The neglected social dimensions to a vehicle-to-grid (V2G) transition: A critical and systematic review," *Environ. Res. Lett.*, vol. 13, no. 1, Jan. 2018, Art. no. 013001.

[3] H. Zhang, Z. Hu, Z. Xu, and Y. Song, "Evaluation of achievable vehicle-to-grid capacity using aggregate PEV model," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 784–794, Jan. 2017, doi: 10.1109/TPWRS.2016.2561296.

[4] O. Sundstrom and C. Binding, "Flexible charging optimization for electric vehicles considering distribution grid constraints," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 26–37, Mar. 2012, doi: 10.1109/TSG.2011.2168431.

[5] D. F. Recalde Melo, A. Trippe, H. B. Gooi, and T. Massier, "Robust electric vehicle aggregation for ancillary service provision considering battery aging," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1728–1738, May 2018, doi: 10.1109/TSG.2016.2598851.

[6] S. Vasantharathna, "Electric power systems," *Electr. Renew. Energy Syst.*, vol. 18, no. 6, pp. 452–455, Jul. 2016.

[7] *Power System Master Plan 2016*, Power Division, Dhaka, Bangladesh, 2016.

[8] D. Toquica, P. M. De Oliveira-De Jesus, and A. I. Cadena, "Power market equilibrium considering an ev storage aggregator exposed to marginal prices—A bilevel optimization approach," *J. Energy Storage*, vol. 28, Apr. 2020, Art. no. 101267, doi: 10.1016/j.est.2020.101267.

[9] S. M. Bagher Sadati, J. Moshtagh, M. Shafie-khah, A. Rastgou, and J. P. S. Catalão, "Operational scheduling of a smart distribution system considering electric vehicles parking lot: A bi-level approach," *Int. J. Electr. Power Energy Syst.*, vol. 105, pp. 159–178, Feb. 2019, doi: 10.1016/j.ijepes.2018.08.021.

[10] Z. Li, Q. Guo, H. Sun, S. Xin, and J. Wang, "A new real-time smart-charging method considering expected electric vehicle fleet connections," *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 3114–3115, Nov. 2014, doi: 10.1109/TPWRS.2014.2311954.

[11] S. I. Vagropoulos, D. K. Kyriazidis, and A. G. Bakirtzis, "Real-time charging management framework for electric vehicle aggregators in a market environment," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 948–957, Mar. 2016, doi: 10.1109/TSG.2015.2421299.

[12] 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.

[13] H. Yi, Q. Lin, and M. Chen, "Balancing Cost and Dissatisfaction in Online EV Charging under Real-time Pricing," in *Proc. IEEE INFOCOM*, Apr. 2019, pp. 1801–1809, doi: 10.1109/INFOCOM.2019.8737558.

[14] Z. Liu, Q. Wu, K. Ma, M. Shahidehpour, Y. Xue, and S. Huang, "Two-stage optimal scheduling of electric vehicle charging based on transactive control," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2948–2958, May 2019, doi: 10.1109/TSG.2018.2815593.

[15] B. Hashemi, M. Shahabi, and P. Teimourzadeh-Baboli, "Stochastic-based optimal charging strategy for plug-in electric vehicles aggregator under incentive and regulatory policies of DSO," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3234–3245, Apr. 2019, doi: 10.1109/TVT.2019.2900931.

[16] G. R. Aghajani, H. A. Shayanfar, and H. Shayeghi, "Demand side management in a smart micro-grid in the presence of renewable generation and demand response," *Energy*, vol. 126, pp. 622–637, May 2017, doi: 10.1016/j.energy.2017.03.051.
[17] S. M. B. Sadati, J. Moshtagh, M. Shafie-khah, and J. P. S. Catalão, "Smart distribution system operational scheduling considering electric vehicle parking lot and demand response programs," *Electr. Power Syst. Res.*, vol. 160, pp. 404–418, Jul. 2018, doi: 10.1016/j.epsr.2018.02.019.
[18] L. Gong, W. Cao, K. Liu, and J. Zhao, "Optimal charging strategy for electric vehicles in residential charging station under dynamic spike pricing policy," *Sustain. Cities Soc.*, vol. 63, Dec. 2020, Art. no. 102474, doi: 10.1016/j.scs.2020.102474.
[19] C. M. García Mazo, Y. Olaya, and S. Botero Botero, "Investment in renewable energy considering game theory and wind-hydro diversification," *Energy Strategy Rev.*, vol. 28, Mar. 2020, Art. no. 100447.
[20] Z. Liu, Q. Wu, S. Huang, L. Wang, M. Shahidehpour, and Y. Xue, "Optimal day-ahead charging scheduling of electric vehicles through an aggregative game model," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5173–5184, Sep. 2018, doi: 10.1109/TSG.2017.2682340.
[21] A. Tiguercha, A. A. Ladjici, and M. Boudour, "Competitive co-evolutionary approach to stochastic modeling in deregulated electricity market," in *Proc. IEEE Int. Energy Conf. (ENERGYCON)*, May 2014, pp. 514–519, doi: 10.1109/ENERGYCON.2014.6850475.
[22] A. Tiguercha, A. A. Ladjici, and M. Boudour, "Suppliers' optimal biding strategies in day-ahead electricity market using competitive coevolutionary algorithms," in *Proc. 3rd Int. Conf. Syst. Control*, Oct. 2013, pp. 821–826, doi: 10.1109/ICoSC.2013.6750952.
[23] J. Vijaya Kumar, D. M. Vinod Kumar, and K. Edukondalu, "Strategic bidding using fuzzy adaptive gravitational search algorithm in a pool based electricity market," *Appl. Soft Comput.*, vol. 13, no. 5, pp. 2445–2455, May 2013, doi: 10.1016/j.asoc.2012.12.003.
[24] P. You, Z. Yang, M.-Y. Chow, and Y. Sun, "Optimal cooperative charging strategy for a smart charging station of electric vehicles," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 2946–2956, Jul. 2016, doi: 10.1109/TPWRS.2015.2477372.
[25] T. Lillicrap, J. Hunt, A. Pritzel, N. Heess, and T. Erez, "Continuous control with deep reinforcement learning," in *Proc. 4th Int. Conf. Learn. Represent.*, 2016, pp. 1–14.
[26] Z. Ding, Y. Huang, H. Yuan, and H. Dong, "Introduction to reinforcement learning," in *Deep Reinforcement Learning: Fundamentals, Research and Applications*. Singapore: Springer, 2020.
[27] G. Barth-Maron, M. W. Hoffman, D. Budden, W. Dabney, D. Horgan, D. Tb, A. Muldal, N. Heess, and T. Lillicrap, "Distributed distributional deterministic policy gradients," in *Proc. 6th Int. Conf. Learn. Represent.*, 2018, pp. 1–16.
[28] R. S. Sutton and A. G. Barto, "Reinforcement learning: An introduction," *IEEE Trans. Neural Netw.*, vol. 9, no. 5, p. 1054, Sep. 1998, doi: 10.1109/tnn.1998.712192.
[29] A. M. Laila, M. A. Sohair, A. H. Mohamed, Y. M., S., A., M., T., S., and A., F., "Value and limitations of polymerase chain reaction and different serological markers in diagnosis of single and mixed acute viral hepatitis," *Assiut. Med. J.*, vol. 27, pp. 1–16, 2003.

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.

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.

LINGXIANG WANG received the B.S. degree in business administration from North China Electric Power University (NCEPU), China, in 2018, where she is currently pursuing the master's degree. Her research interests include power market analysis and power load management.

HEPING JIA received the B.E. degree in electrical engineering from North China Electric Power University (NCEPU), China, in 2014, and the Ph.D. degree in electrical engineering from Zhejiang University, China, in 2019. Since July 2019, she has been a Postdoctoral Researcher with NCEPU. Her research interests include reliability assessment of power systems and risk analysis of engineering systems.

MENGSHU SHI received the master's degree from NCEPU, in 2020, where she is currently pursuing the Ph.D. degree with the School of Economics and Management. Her main research interests include low-carbon and energy economy development and comprehensive energy systems.