A DQN-based Two-Stage Scheduling Method for Real-Time Large-Scale EVs Charging Service

Tianyang Li[1], Yingnan Han[2], and Xiaolong Li[3*]
[1,2] School of Computer Science, Northeast Electric Power University
Jilin, Jilin, 132012 China
[e-mail: 872888861@qq.com, 2061331823@qq.com]
[3] Harbin Branch, Bank of Inner Mongolia Co. Ltd
Harbin, Heilongjiang, 150000 China
[e-mail: long1378568805@163.com]
[*]Corresponding author: Xiaolong Li

Abstract

With the rapid development of electric vehicles (EVs) industry, EV charging service becomes more and more important. Especially, in the case of suddenly drop of air temperature or open holidays that large-scale EVs seeking for charging devices (CDs) in a short time. In such scenario, inefficient EV charging scheduling algorithm might lead to a bad service quality, for example, long queueing times for EVs and unreasonable idling time for charging devices. To deal with this issue, this paper propose a Deep-Q-Network (DQN) based two-stage scheduling method for the large-scale EVs charging service. Fine-grained states with two delicate neural networks are proposed to optimize the sequencing of EVs and charging station (CS) arrangement. Two efficient algorithms are presented to obtain the optimal EVs charging scheduling scheme for large-scale EVs charging demand. Three case studies show the superiority of our proposal, in terms of a high service quality (minimized average queuing time of EVs and maximized charging performance at both EV and CS sides) and achieve greater scheduling efficiency. The code and data are available at THE CODE AND DATA.

1. Nomenclature

EV_i	the ith EV (electric vehicle)
CS_j	the jth CS (charging station)
$CD_{(k,j)}$	the kth CD (charging device) of CS_j
$ot_{(k,j)}$	the occupied time for $CD_{(k,j)}$
qt_{ijk}	the queueing time of EV_i at $CD_{(k,j)}$
it_{jk}	the idling time of $CD_{(k,j)}$
AQT	the list contains the average queueing time of EVs in each CS
AIT	the list contains the average idling time of EVs in each CS
CNS	the list contains the EV amount queueing at each CS
N_{ev}	the total number of currently schedulable EV
N_{cs}	the total number of available CS

2. Introduction

$\mathbf{T}$he EVs have increased rapidly with the advantage of environmentally friendly and cost saving [1]. However, lack of charging infrastructure is one of the roadblocks for promoting the penetration of EV [2]. In some scenarios, such as extremely cold weather and holidays, it is a common phenomenon that the large-scale EVs charging request increasing suddenly [3]. The unreasonable EVs charging scheduling strategy might cause a long queueing time [4]. Therefore, dealing with the large-scale EVs charging scheduling requests properly is still a critical question need to be solved for improving the quality of EV charging service (minimizing average queuing time of EVs and maximizing charging performance at both EV and CS sides).

Currently, many studies are conducted for solving the EVs charging scheduling problem. Some of them introduced the global aggregator (GA) based scheduling strategies [5]. In study [6], for example, a GA-based pre-empted EVs charging position reserving algorithm was proposed for a better quality of EV charging service. The GA, as a third-party to cooperate with CSs and EVs, offers scheduling information from a global perspective under large-scale EV charging requests scenarios. Some studies have proposed the heuristic algorithm-based EV charging scheduling method to deal with the large-scale surging requests according to the information offered by GA, e.g., genetic algorithm [7-9], PSO (particle swarm optimization) [10-12], artificial bee colony [13]. However, these methods are easily trapped into local optimization and need more time to get a suboptimal solution under a large solution space of problems with surge charging demand.

For avoiding the drawbacks and generating scheduling plan according to the dynamically changing environment, other studies were devoted to exploring different scheduling approaches based on reinforcement learning (RL). Quite a few researchers use reinforcement learning (RL), or deep reinforcement learning (DRL) based algorithm to solve the problem of EV charging scheduling from different perspective. Some studies apply DRL-based algorithms for minimizing charging cost of EV users [14], increasing profit of charging service providers [15], or balancing electric load profiles [16, 17]. Obviously, the previous studies approached their optimizing target based on the historical data, i.e., time series, it difficult to cope with the surging EV charging scheduling problem due to the models is not easily to be applied in the situation with large-scale action space and high real-time demand. For coupling with the large -scale charging demands and getting a better quality of service (the low average

queuing time of EV and low average idling time of charging devices), existing RL based algorithms have common difficulties, mainly including insufficient states, unreasonable designs of RL models as well as lacking effective learning algorithms, in solving problems with a high-dimensional action space and state space. For overcoming these difficulties, creating a suitable RL based method for surge charging demands must be undertaken firstly. The definition for state of this method should be characterized by different levels of service qualities and may consider both EVs and CSs sides. Meanwhile, designs of RL model should be based on scheduling processes. In addition, training algorithms should pay great attention to details of scheduling strategies while avoiding trapped into local optimal solution.

Despite the importance of providing a practical RL based method for the large-scale EVs charging scheduling problem from a global perspective, the current literature only refers to theoretical methods to gain strategic, marketing, or operational insights. Currently there is no comprehensive RL based models that solves the large-scale EVs charging scheduling in the literature. The lack of literature precedent raises three questions:

1. How to define fine-grained states that consider both EVs and CSs sides?

2. How to design RL models with scheduling processes?

3. What is the key to gain the optimal solution with a high-dimensional action space without trapping into local optimal solution?

For solving these questions, we propose an innovative DQN-based two stage scheduling method for enhancing the service quality of large-scale EVs charging scheduling. Fine-grained states that describe both EVs and CSs sides are defended. Two networks based on the scheduling process are designed and fine-tunning algorithms are presented to improve the effectiveness and efficiency. With four well-designed experiments, our study is testified that could generate a better charging scheduling plan under surge charging scheduling demands. Furthermore, the proposed deep reinforcement learning based technology roadmap can be extended to other problems with large action space and complex state representation.

The rest of this paper is organized as follows. "Related Work" investigates the research results related to EV charging scheduling problems. "Problem Definition" describes the concepts involved and introduces the mathematical model of optimization objective. "DQN-based Two-stage Scheduling Method" section details proposed method. Finally, the "Case Studies" section compares and analyzes the performance of the proposed method with the current major algorithms and proves its superiority. "Conclusion" concludes this algorithm and gives an outlook on possible future research directions.

3. Related Work

Using RL based methods to solve a problem that arranges each EV to an appropriate charging station is difficult because of the reward setting, complex states defining, and models designing. While many studies still focus on employing RL to solve EVs charging scheduling due to its ability that could gain an optimal action strategy from the action space [18].

The reasonable reward is of great importance for gaining the target scheduling goal. The designing of reward is varying for different research with different EVs charging scheduling targets. The reward can be the queueing time minimizing, charging efficiency improvement by reducing average idling time of charging devices and get more profits for charging stations [19, 20].

Generally, the designing of state depends on scheduling goals. States used in studies include the waiting time [21], charging price [20], traffic condition, battery SOC [22], EVs locations [19], charging price [23], degrees of user satisfaction and etc. For increasing individual

satisfaction, energy consumption levels or the expected charging duration are also considered when designing state representations. However, it is still lacking appropriate state representations for the large-scale EVs charging scheduling problem to gain a better charging effectiveness and efficiency [24].

Based on warily designed states, an appropriate action for selecting EVs to be charged or scheduling an EV to a CS can be decided by the models. The form of a model of RL is a single deep neural network or combination of some interactable neural networks. For some scenarios with simple discrete action spaces, a model can be a table that contains state-action pairs with their values [25]. However, this kind of table is not suitable for real world applications with a large state space and high-dimensional action space. Some studies proposed more sophisticated models, such as combining DDPG with DQN, to predict the action value after extracting state features by LSTM due to the high complexity of state-action mapping [26]. Nevertheless, these sophisticated models also cannot gain the optimal solution efficiently due to the large-scale of mapping spaces. Decomposition the action space and design suitable model based on the scheduling process might be a promising way to deal with the large-scale action space. While existing methods pay little attention to this.

In summary, most aforementioned studies have only dealt with charging scheduling problem with the small-scale of EVs and solved to the low-dimensional state spaces. Few studies focus on the large-scale EVs charging scheduling problem for decreasing queueing time and idling time for both EVs and CSs to improve the service quality. Therefore, an efficiency scheduling method for the surge large-scale EVs scenarios is needed. For solving this problem, the RL based method should be proposed with:

1) fine-grained states that consider both EVs and CSs sides.

2) global optimizing target for improving QoS (minimizing average queuing time of EVs and maximizing charging performance at both EV and CS sides) with the suitable DRL-models.

3) action selecting mechanism with fine-tunning algorithms to generate appropriate EV scheduling strategy.

4. Problem Definition

Our proposal focuses on offering a charging scheduling strategy with high charging service quality (minimized average queuing time of EVs and maximized charging performance at both EV and CS sides) for the large-scale EVs. For describing the proposed method clearly, the mathematical background is shown as below.

4.1 Entity Definition

The entities that in proposed algorithm are EV, CS, CD and GA. As a CPS (cyber physical system), global aggregator is a third-party platform for offering real-time traffic, EV, and CS information according to the interfaces of EV industry, charging service providers and map companies. The definition of EV, CS and CD are summarized as **Table 1**.

Table 1. Definition of entities

Item	Definition
EV_i	$\langle i, elo_i, ltd_i, ct_i \rangle$
CS_j	$\langle j, clo_j, dn_j \rangle$
$CD_{(k, j)}$	$\langle j, k, ot_{(k, j)} \rangle$

4.2 Assumption

The large-scale EVs charging scheduling request occurs in a city scenario with a complex traffic situation. Without the loss of generality, the location of EVs make the charging request is randomly around the charging stations. At same time, the total amount of EVs is not predefined as well as K initial CSs are available. All the entity defined in 4.1 can communicate with GA in real-time. In addition, the priority of EVs is not pre-defined and all details of the scheduling plan is generated by algorithms.

4.3 Mathematical Model

Our proposal focuses on offering a charging scheduling plan with a high service quality for the large-scale EVs. To describe the proposed method clearly, the mathematical background is shown as below.

The problem is defined in a discrete action space and has a discrete time setting. For the given EV_i, CS_j, the travel time from EV_i to CS_j is defined as (1):

$$tt_{ij} = \frac{d_{(i,j)}}{\alpha \times v_i} \tag{1}$$

where $\alpha \in [0, 1]$ is a factor shows the congestion level, it is determined according to the real-time road congestion level and acquired from the service interface of map companies. The α is bigger when the traffic situation is better otherwise the traffic situation is worse. When EV_i is scheduled at $CD_{(k,j)}$, the expected queueing time of EV_i and the idling time of $CD_{(k,j)}$ can be updated as (2) and (3):

$$qt_{ijk} = \begin{cases} ot_{(k,j)} - tt_{ij} & ,if \ ot_{(k,j)} > tt_{ij} \\ 0 & ,if \ ot_{(k,j)} \leq tt_{ij} \end{cases} \tag{2}$$

$$it_{ijk} = \begin{cases} tt_{ij} - ot_{(k,j)} & ,if \ tt_{ij} > ot_{(k,j)} \\ 0 & ,if \ ot_{(k,j)} \geq tt_{ij} \end{cases} \tag{3}$$

After a charging device $CD_{(k,j)}$ allocated to a EV_i, the occupied time $ot_{(k,j)}$ can be update as (4):

$$ot_{jk} = \begin{cases} tt_{ij} - ct_i & ,if \ ot_{jk} \leq tt_{ij} \\ ot_{jk} + ct_i & ,if \ ot_{jk} > tt_{ij} \end{cases} \tag{4}$$

The goal of our proposal is to find a scheduling plan with minimized average queueing time of EVs and average idling time of CSs. Because of the complexity of the goal (optimizing the average queueing time and average idling time at same time), we convert the optimizing object into a single-object optimization problem. The optimization objective can be described as (5):

$$\begin{cases} \min \ \omega_1 \times \sum_k \sum_i \sum_j qt_{ijk} + \omega_2 \times \sum_j \sum_k it_{jk} \\ \\ s.t. \begin{cases} d_{(i,j)} < ltd_i \\ \sum_n \omega_n = 1 \end{cases} \end{cases} \tag{5}$$

5. DQN-based Two-stage Scheduling Method

The traditional DQN-algorithm have barriers on convergence speed when solving problems with a high-dimensional action space. To deal with this problem, a fine-grained state

representation and two independent neural networks were designed to select an appropriate CS for EV.

5.1 State representation

State designing is of great importance for DRL. From the perspective of the optimization target of the large-scale EVs charging scheduling problem, two kinds of states from both EV and CS sides were designed to represent the global information of EVs and CSs.

For any schedulable EV_i, its feature SE_i can be formed as follows:

$$SE_i = [dis\_to\_cs(i), ct_i, ltd_i]^\top \tag{6}$$

Let L denotes the dimension of SE_i, then the feature of charging stations SC (which shape is $L \times 4$) can be depicted as follows:

$$SC = [AQT, AIT, CNS, occup\_T]^\top \tag{7}$$

The global state representation of all schedulable EVs can be formed as follows by combining with the SC, SE_i:

$$SEV = \begin{bmatrix} SE_1 & \cdots & SE_N & SC \end{bmatrix}^\top_{L \times (N_{ev} \times 4)} \tag{8}$$

After an EV numbered with k is selected as the optimal selection, the state for EV_k to select an appropriate charging station which can be describe as follows:

$$SCS = \begin{bmatrix} SE_k \\ SC \end{bmatrix}_{(5 \times L)} \tag{9}$$

5.2 Structure of Neural Network

<div align="center">Table 2. Structure of proposed neural network</div>

Parts	Layers	Adopt by	Layers Abbreviation	Input data shape
The EV Selecting Model	conv layer with kernel size = (8×8)	*relu*	EC1	$(N_{ev}+4) \times L$
	conv layer with kernel size = (4×4)	*relu*	EC2	-
	conv layer with kernel size = (2×2)	*relu*	EC3	-
	fully connected layer with nodes = K_e	*relu6*	EF1	K_e
	fully connected layer with nodes = K_e	*relu6*	EF2	K_e
	fully connected layer with nodes = N_{ev}	*relu6*	EF3	N_{ev}
	fully connected layer with nodes = N_{ev}	*softmax*	EF4	N_{ev}
The CS Selecting Model	conv layer with kernel size = (3×6)	*relu*	CC1	$5 \times L$
	conv layer with kernel size = (2×4)	*relu*	CC2	-
	conv layer with kernel size = (2×4)	*relu*	CC3	-
	fully connected layer with nodes = K_c	*relu*	CF1	K_c
	fully connected layer with nodes = K_c	*relu*	CF2	K_c
	fully connected layer with nodes = N_{cs}	*relu*	CF3	N_{cs}

For solving the problem on high dimension (in our proposal, we found that fewer studies have been conducted specifically for the scenario where the number of EVs is much larger than the number of available charging stations. Based on this fact, the charging scheduling problem when the amount of the EVs is larger than the one-time simultaneous charging capacity, i.e., the number of charging device, of all the charging stations is considered as a large-scale EV scheduling problem in this study. Relevant details about the data are presented later in the experimental stage) of action space for the large-scale EVs charging scheduling, the two interactable neural networks were designed for the EVs selecting and CS assignment. The

models are designed to mapping the state-action pair to a Q-value which could help EV/CS selecting. As shown in the **Table 2**, the state *SEV* is input into the convolutional layer (*EC1*, *EC2*, *EC3*) where each convolutional layer was adopted by a rectified linear unit (*relu*). A feature map (with K_e features) extracted by convolutional layer is then fed into the fully connect layers (*EF1*, *EF2*, *EF3*) to select an optimal EV. Therein, the size of fully connected layer depends on the real problem. After an optimal EV_k was selected, *SCS* is constructed as an input of convolutional layer *CC1*. Then, the appropriate charging station can be selected by the Q-values for CS-selection acquired from fully connected layer (*CF1*, *CF2*, *CF3*) according to the feature map (with K_c features) generated by conv layer (*CC1*, *CC2*, *CC3*). For improving the ability that fitting the Q-function for EV and CS, the *relu6* functions are added into each two adjacent fully connect layers.

Both neural networks for selecting EVs and the corresponding appropriate charging stations have a fully connected layer part, and the number of network nodes is designed to consider the model's feature extraction of the input states as well as the number of objects (EVs or CSs) to be selected in the actual application scenario. Considering the complexity of the problem, although a larger number of nodes in the intermediate layer would result in a better fit of the model to the problem, the size of the nodes in the intermediate part of the fully connected layer was limited as the same nodes with the dimension of input feature in order to make the model more generalizable, and its validity was fully verified in the experiments.

5.3 Training Algorithm
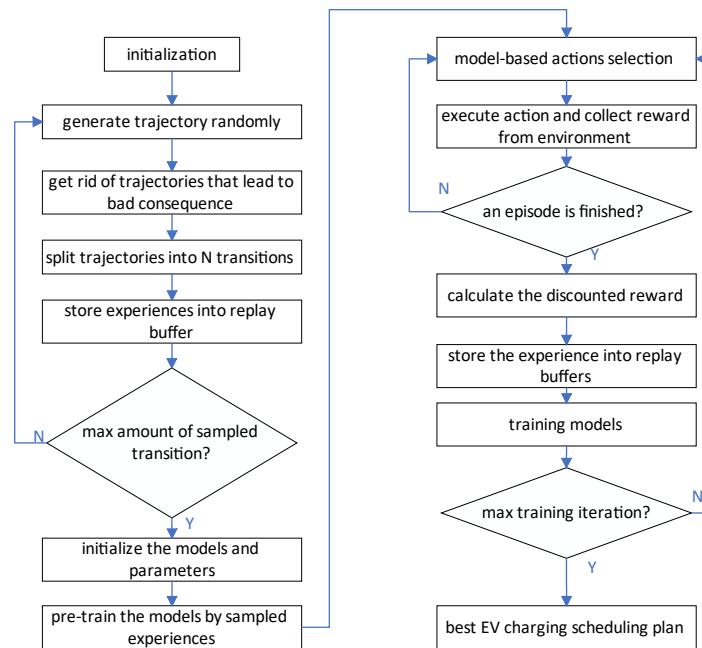
Fig. 1. The flowchart of our algorithm

With the two models (the EV selecting model and the CS selecting model), an EV can be selected according to the Q-value of EV (weighted value of average queueing time and average

idling time on currently scheduled EV discounted by γ) predicted by the EV-selecting network and an appropriate CS for EV can be selected according to the Q-value of CS (expected queueing time of EV at target CS) predicted by the CS-selecting network. The update process can be conducted according to the difference between the output of network for CS (or EV) and the desired Q-value (weighted Q-value) returned from the environment.

As shown in the **Fig. 1**, the training of proposed model has two main process includes pretraining stage and training stage. Firstly, the models are pre-trained based on well-performed experience collected and sampled from interacting with environment randomly. Secondly, a training process to explore the solution space by linear reducing ε-greedy method was introduced for a better performance on EV charging scheduling.

For the proposed models (EV-model and CS-model), the pre-train stage collect experience from interacting with environment for *IS* (a constant) times to train EV-model and CS-model separately. Then, the EV-model was trained by experience stored in *ERB* (the replay buffer for storing the experiences to train the EV-model) for *PTS* times and CS-model was trained by experience stored in *CRB* (the replay buffer for storing the experiences to train the CS-model) for *TS* times. The pseudo-code for pre-training and training the models is summarized as Algorithm 1.

Algorithm 1: The training of the proposed models
INPUT: γ, *IS*, *TS*, *PTS*, *ERB*, *CRB*, un-trained models
OUTPUT: the model that can generate the best scheduling plan
PRE-TRAINING STAGE
1   **LOOP** *random-step* from 0 to *IS*:
1.   generate a trajectory randomly
2.   **IF** all episodes are finished collecting:
4.   filter out the first 30% of trajectories with lower *AQT*
5.   store experience into *ERB* and *CRB*
6   **LOOP** *pre-training-step* from 0 to *PTS*:
7.   pre-training the model
TRAINING STAGE
8.   **LOOP** *i* from 0 to *TS*:
9.   **LOOP** *step* from 0 to N_{ev}:
10.   select action a_{et} and a_{ct} for EV and CS according to Algorithm 2 separately. at time step *t*
11.   execute the actions
12.   obtain the reward and next state from environment for EV and CS separately
13.   store the experience into *ERB* and *CRB* separately
14.   **LOOP** *training-step* from 0 to *TS*:
15.   training the model

Generally, selecting EV and CS with the highest Q-value in every step might lead to a bad scheduling performance while the models are not convergence. It might introduce worse experience into replay buffer. For solving this problem, a ε-greedy strategy with linear decreased ε is introduced. To avoid introducing poor action selection by ε-greedy strategy, a local queuing order fine-tunning scheme is introduced to offset the randomness brought by the original exploration method. Let $qn_{(k, j)}$ denotes the number of EVs, the Algorithm 2 shows the action selecting algorithm with fine tuning.

Algorithm 2: The fine tuning of scheduling consequence
INPUT: EV state, CS state, $\varepsilon \in [0, 1]$, EV model, CS model, random variable e, c between 0 and 1
OUTPUT: optimal selection for EV and CS
DEFINE: $M_i = dis_{(j, k)} + CT_i$ (if EV_i scheduled at $CD_{(j, k)}$)
1. calculate Q-values for all not-scheduled EVs according to EV-state
2. **IF** $e < \varepsilon$:

<div align="right">**Algorithm 2**</div>

3.   select EV randomly
4. **ELSE**:
5.   select the EV with optimal Q-value
6. construct CS-state
7. calculate Q-value for all CSs according to CS-state based on the CS-selecting model
8. **IF** $e < \varepsilon$:
9.   select CS randomly
10. **ELSE**:
11.   select CS with optimal Q-value
12. scheduling the selected EV to the selected CS
FINE-TUNNING STGAE
13. **IF** all EVs are scheduled:
14.   **FOR** j in range 0…N_{cs}:
16.   **FOR** k in range 0…number of available charging device of CS_j:
17.   move EVs at device k of CS_j to set S and calculate M value for each EV
18.   **FOR** i in range 0… (the number of EVs scheduled at $CD_{(k, j)}$):
19.   select EV_i with minimum M_i and schedule EV_i to $CD_{(k, j)}$
20.   remove EV_i from S and update M-$value$ for each EV in S

6. Case Studies

6.1 Yardstick and Dataset

All experiments are implemented by using PyCharm Integrated Development Environment version 2019.1.1 with Python 3.7 on a PC: 11th Gen Intel(R) Core (TM) i7-11800H @ 2.30GHz with 32G memory and Nvidia RTX 3060 with 6G memory. To investigate the performance of the proposed method, three case studies are designed based on data collected from a real-world dataset from a private EVs charging service. In our experiments, a relatively bad traffic situation is considered, the congestion level parameter α is set to be 1.

The dataset (described in **Table 3**) can be downloaded from (data source). It can be divided into three parts, which are the EV information, CS information and charging price information. The EV information includes all of the details for EVs, which are the initial coordination, the left traveling distance, and the expected charging time of each EVs. The details about CS information includes the coordination of each available charging stations and the charging price per hour.

Item	Value
The number of CSs	34
The charging device of each CS	5
The charging voltage of charging device	500V-700V
The total number of EVs that are intending to use the charging device in a certain duration	1000
The number of EVs that can arrive the nearest CS with its remaining battery power	899
The number of EVs that use the batteries exchanging service	101

The entities in the experiments are 1000 EVs and 34 CSs. Each CS has 5 high voltage available charging devices and the EVs are randomly located around the charging stations. In our research, 101 EVs cannot be scheduled due to the extremely low battery, which need to use the battery exchanging service. Only 899 EVs could reach at least one charging station with their remaining capacity of batteries.

6.2 Experiment Setting

We designed three case studies for fully analyzing our proposal. Considering the fact that the amount of CS is 34, the L is set to be 34+2 (therein, the "34+2" denotes the distance between EV_i and all CSs (34 CSs), expected charging time and left traveling distance).

In case study I, the effect of reward designing on charging scheduling quality was investigated.

In case study II, we compared the ability on solution space exploring between the proposed algorithm and EDA-GA based EVs scheduling algorithm with 200 and 625 iterations. The superiority of scheduling order fine-tuning is proved by ablation experiments. And the lower learning rate is better for searching optimal scheduling scheme is then discussed.

In case study III, the proposed large-scale EVs charging scheduling algorithm is compared to the algorithms in the literature. Four groups of experiments were conducted. Each group of experiments are described as follows.

The first group contains the experiments of customer-oriented FCFS algorithm where the customers select the nearest CS and supplier-oriented greedy scheduling algorithm which is maximizing supplier revenue.

The second group contains genetic algorithm-based random FCFS algorithm, genetic algorithm based random scheduling algorithm and genetic algorithm-based greedy EV charging scheduling algorithm with relative low queueing time and idling time.

The third group contains EDA-GA based genetic algorithm under 200 and 625 iterations.

The last group of experiment contains DRL-based algorithms including traditional DQN, DDPG, A2C, and CDDPG.

6.3 Case Study I

This section, A group of the control experiments based on different reward settings are conducted. The experiments investigate the effect of the designing for reward on the quality of EV charging scheduling service.

**Table 4.** Scheduling results based on different rewards setting

| EV-Target	CS-Target	best queueing time	best idling time
$0.5\times RAW+0.5\times MNCSD$ (3)	$csval+1.5\times CEN_k$	26.330	7.087
$0.5\times RAW+0.5\times MNCSD$ (5)	$csval+1.5\times CEN_k$	42.273	4.357
$0.5\times RAW+0.5\times MNCSD$ (2)	$csval+1.5\times CEN_k$	25.779	7.148
$0.5\times RAW+0.5\times MNCSD$ (2)	$csval+2.0\times CEN_k$	27.901	8.330
$0.5\times RAW+0.5\times MNCSD$(3)	**$csval+2.5\times CEN_k$**	**21.300**	**4.930**

The different reward settings and the scheduling consequences are shown in **Table 4**, where the $MNCSD(k)$ is the average distance from an EV to its nearest k charging station, the RAW is a weighted value of sum(AQT)/N_{ev} (sum(AQT) denoted the summation of queueing time for all EVs) and AIT/N_{ev} (sum(AIT) denoted the summation of idling time for all EVs) of EVs (in our experiment setting, the weight of sum(AQT)/N_{ev} is 0.8 and the weight of sum(AIT)/N_{ev} is 0.2), the CEN_k used to represent the congest degree of charging station CS_k and the $csval$ is the Q-value predicted by CS-selecting network. A conclusion can be drawn from **Table 4** that the quality of large-scale EVs charging scheduling service is affected by expected queueing and crowding degree of target charging station. It can obtain a better charging service if an EV is scheduled at a charging station with low congestion level. On another hand, it could obtain a shorter average queueing time and average idling time when an EV with the shorter distance to nearby charging station is scheduled firstly.

6.4 Case Study II

To show the superiority on EV charging scheduling performance of our proposal, a group of comparing experiments are conducted.

Firstly, the advantage of scheduling consequence fine tuning method that could avoid local optimal solutions is proved.
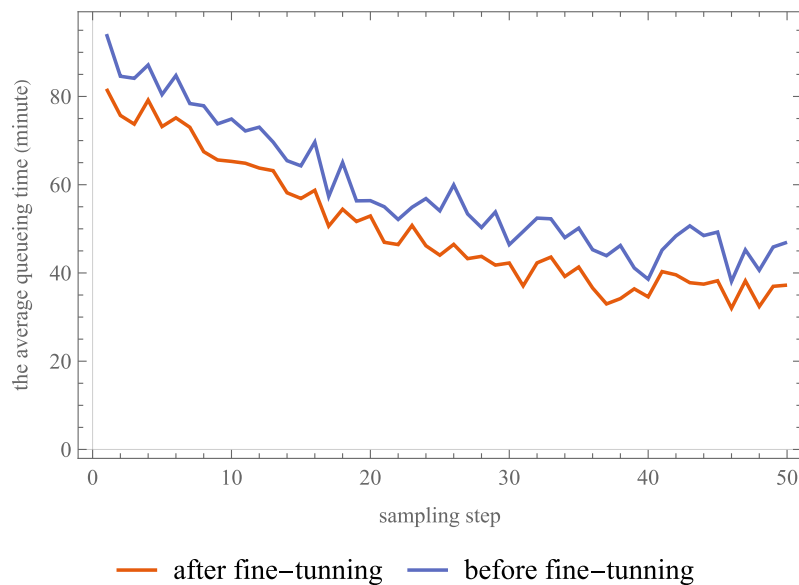
**Fig. 2.** The comparison of *AQT*
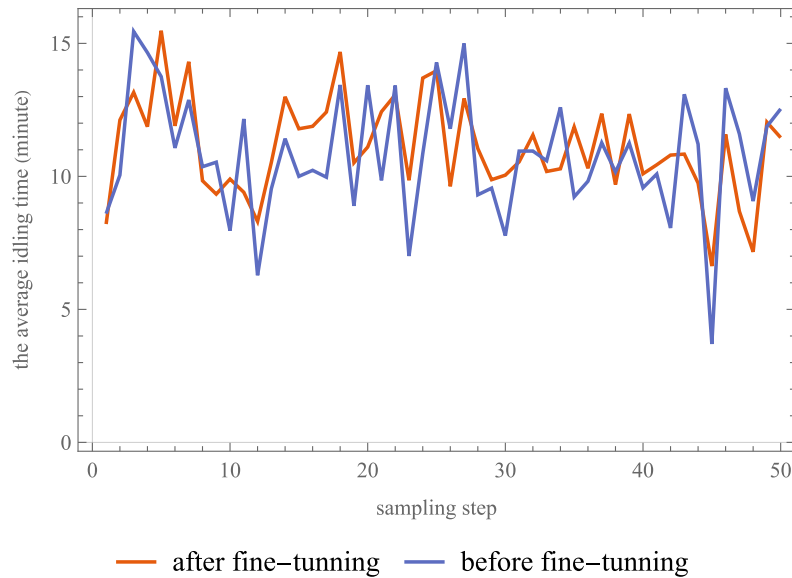
Fig. 3. The comparison of *AIT*

Secondly, the comparison of the scheduling consequences between other well-performed method and our proposal is conducted. The consequences show the effectiveness of our proposal on action space exploring.

For investigating the advantages of scheduling consequences fine-tunning method, the ablation experiments are conducted. **Fig. 2** and **Fig. 3** show the comparison of *AQT* and *AIT* before and after the fine tuning.

As we can conclude from the figures, the scheduling consequences after fine-tuning is relatively better than before. With the *AQT* of all EV decreasing, the *AIT* for charging stations also shows a downward trend. At the same time, it can also be seen that while the average queue time gradually decreases, the average idle time shows frequent fluctuations. This indicates that the queueing order corresponding to the electric vehicles that queuing in front of each charging device is constantly changing in the exploring stage. Furthermore, the experimental results can also demonstrate that the control group with fine-tuning algorithms has stronger ability in generating the optimal solution for electric vehicles. This fully proves that our proposed queuing order fine-tuning method has played a promoting role in the optimization process of the model in the solution space.
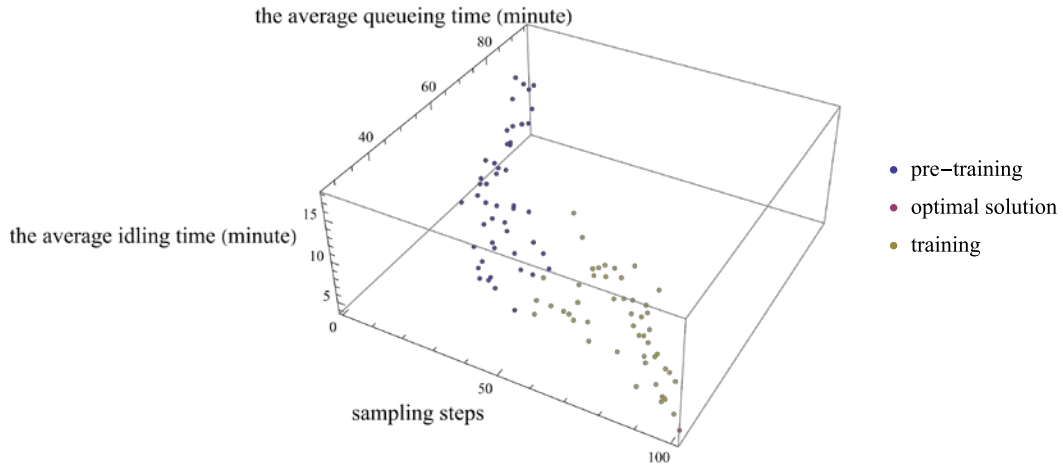
**Fig. 4.** The exploring of solution space in our proposal

Based on the scheduling consequences fine-tuning, the overall EV scheduling performance in solution space exploration of our proposal is shown in the **Fig. 4**. As shown in the figure, the proposed algorithm can be divided in to two main parts which are the pre-training stage (blue dots) and the solution space exploring stage (yellow dots). At the beginning of the pre-training stage, the solutions are relatively scattered in the left-top of the figure because of the higher ε and not converged models. With the iteration increasing, the more well-performed experiences are collected and used to train the models and the founded solutions began to concentrate (the part with queuing time around 30. However, there is few solutions (blue point) acquired in pre-train stage have a higher performance which might be due to the randomness of ε-greedy method. At the training stage, it is clearly that the searching ability to find better solution is enhancing as the iteration step increasing. At the beginning of the training stage, it is relatively slow (it takes around 30 sampling steps) to find a better solution due to the higher ε. With the linear ε decreasing strategy, the scheduling plan with better performance (lower average queuing time for EVs and lower idling time for charging stations) can be generated by the proposed algorithm. In addition, it can be concluded from **Fig. 4** and **Fig. 5** that comparing to the EDA-GA methods, the proposed algorithm has more stable property to exploring the solution space. The better solution can be generated with more iterations. For EDA-GA methods, however, the inner mechanism might be leading to a worse performance on EV charging scheduling with more iterations (as shown in **Fig. 5**).

Fig. 5 illustrates the exploration process of the control group in the solution space. Comparing with our proposal, the control group algorithm (the EDA-GA) can find relatively excellent solutions at the initial stage. For example, the initialized EV charging scheduling strategy given by the control group achieves a near-optimal solution. However, in the process of continuous and deeper exploration, both sets of experiments (EDA-GA with 125 and 625 iterations) in the control group show some stagnant characteristics, i.e., the method does not give a better exploration of the solution space. The relative dispersion of the locations of the solutions found in the solution space by our proposed method compared to the results in **Fig. 5** shows that the EDA-GA based method has some shortcomings in the optimization (exploration of the solution space) capability. This also confirms that our proposal has the characteristic of being able to explore the solution space well.
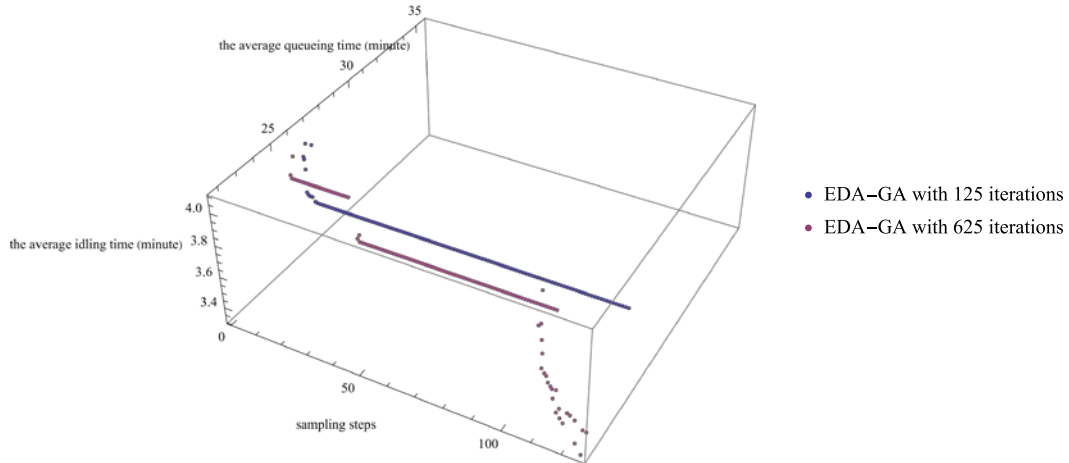
Fig. 5. The exploration of EDA-GA

From **Fig. 6**, it can be concluded that in the early stage of model training, the low learning rate leads to a slow fitting to the action value function, and the efficiency of EV charging scheduling tends to be consistent with its control experiment with learning rate 0.001. However, as the continues to decrease (sampling step 40-50), the difference between the two in terms of scheduling results starts to become larger. We suppose that this is because at the later stage of the exploration, smaller ε tends might introduce a less randomness while selecting an action, and the model training for the predicting of action value is mainly determined by the learning rate. Since the model involves many model parameters, a higher learning rate may, to some extent, result in ignoring some solutions in the solution space that are more suitable for the actual problem, which making the model easy to converge to suboptimal solutions.
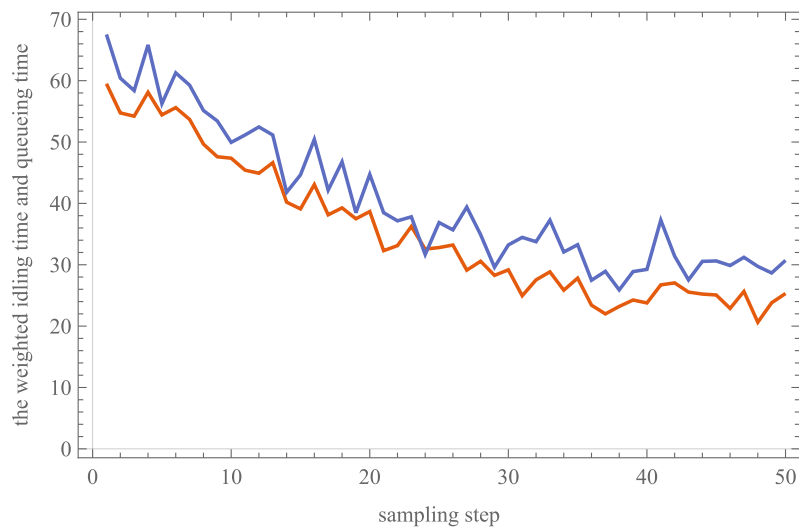
Fig. 6. The comparison of different learning rate

6.5 Case Study III

To describe the superiority of the proposed algorithm, our best scheduling consequence (average queueing time = 21.300 and average idling time = 4.930) were compared to other algorithms with currently well results for EVs charging scheduling under the same large-scale EVs charging scheduling environment. The result is shown as **Table 5**.

Table 5. Scheduling results based on different scheduling algorithms

Classifications	No.	Algorithm	Iteration number	average idling time	average queueing time	time consumed (seconds)
dispatching rule	1	customer oriented FCFS	-	3.00	80.80	0.15
greedy based	2	supplier oriented greedy scheduling	-	1.26	77.10	0.17
	3	genetic based random FCFS	200	12.30	35.10	411.50
genetic based	4	genetic based random scheduling	200	8.90	31.80	513.00
	5	genetic based greedy EVs charging scheduling	200	4.16	25.10	519.9
EDA-based	6	EDA-GA based genetic. algorithm	200	3.00	21.60	1195.0
	7	EDA-GA based genetic. algorithm	625	3.20	21.80	3235.00
(D)RL-based	8	DQN-based EV charging scheduling	-	7.34	24.98	27.14
	9	DDPG based EV charging scheduling	-	4.39	21.78	25.98
	10	A2C based EV charging scheduling	-	3.54	23.12	26.1
	11	**OUR PROPOSAL**	-	**4.93**	**21.30**	**30.20**

Comparing with the EDA-GA based large-scale EVs charging scheduling algorithm with 625 iterations, the average queueing time is declined to 21.3 (about 2.29%). From the point of EVs charging service result, our method can generate the optimized scheduling strategy and at same time decrease the queuing time when EV users waiting for charging in front of charging device. On the other hand, our proposed large-scale EVs charging scheduling algorithm can give a suitable charging scheme in a short time compared to other algorithms

with similar scheduling results. Comparing with the EDA-GA based large-scale EVs charging scheduling algorithm with 125 iterations, the proposed algorithm can generate a charging scheme for surge demand in nearly 30 seconds (66.05 times faster than the compared algorithm). Additionally, we can draw a conclusion by analyzing **Table 5**. that there exists some difference between scheduling plan even though the scheduling service qualities are similar. It can be considered that there might exists more than one optimal scheduling plan with different scheduling order while keeping the same queueing time for the large-scale EVs and the idling time for charging stations due to the high dimensional action space and state space. By comparing the result of experiment 6 and experiment 7, the proposed algorithm can increase idling time while keeping queueing time at a relative low level which has a positive effect on relieving the discharging pressure of the charging station.

On another hand, by comparing the performance of the well-performed DRL-based EV charging scheduling algorithms architectures, such as experiment 8 and experiment 9, in our experiment environment, it can be concluded that our proposal is well-performed on large-scale EV charging scheduling task over the result in experiment 8 than CDDPG-based algorithm. In addition, our proposal enhanced the performance of traditional DQN in complex action space by fine-tuning the EV charging scheduling consequences as well as the well-interacted neural network. It clearly that our proposal enhanced the performance in large-scale EV charging scheduling environments.

In addition, compared to another type of reinforcement learning based method, such as A2C, which is an EV scheduling strategy based on the policy gradient method, our proposal is still slightly superior to such methods. This suggests that among large-scale EV charging scheduling problems, scheduling strategies are required differently at different stages, and adaptive strategies generated from a global perspective may be flawed under complex EV charging scheduling conditions. The Q-value based reinforcement learning method is able to adequately estimate the action state values reasonably in such scenarios, which leads to the reason why our proposal is better than the A2C method.

7. Conclusion

Given the reality that the bad service quality and low execute efficiency of current scheduling methods for the large-scale EVs charging, a DQN based two-stage scheduling method is proposed. Based on the designed fine grained state representation and two delicate neural networks, the ε-greedy strategy was used to effectively explore the action space. For avoiding the negative impact of inner randomness, training algorithms for fine-tunning networks were proposed. Results of comparative experiments have shown that, our method for the large-scale EVs charging scheduling has a positive impact on improving the quality of EVs charging service.

In this work, we proposed a clear roadmap for solving complex problems by using DRL. We designed an architecture for solving continuous decision problem by deep reinforcement learning. Under this architecture, the problem with large state space and action space is solve with the interactable neural networks and warily designed state. In addition, our method can easily expand to other problems with large state action space (such as unrelated parallel machine scheduling, which has the similar scheduling pattern with EV charging scheduling problems).

In future, the target of our research will focus on large-scale EVs charging scheduling problem under a more complex traffic scenario.

References

[1] T. Long, Q.-S. Jia, G. Wang, and Y. Yang, "Efficient real-time EV charging scheduling via ordinal optimization," *IEEE Transactions on Smart Grid*, vol. 12, no. 5, pp. 4029-4038, 2021. Article (CrossRef Link)

[2] Z. Wei, Y. Li, Y. Zhang, and L. Cai, "Intelligent parking garage EV charging scheduling considering battery charging characteristic," *IEEE transactions on industrial electronics*, vol. 65, no. 3, pp. 2806-2816, 2018. Article (CrossRef Link)

[3] T. Li, X. Li, T. He, and Y. Zhang, "An EDA-based Genetic Algorithm for EV Charging Scheduling under Surge Demand," in *Proc. of 2022 IEEE International Conference on Services Computing (SCC)*, pp. 231-238, 2022. Article (CrossRef Link)

[4] M. M. Rahman, E. A. Al-Ammar, H. S. Das, and W. Ko, "Comprehensive impact analysis of electric vehicle charging scheduling on load-duration curve," *Computers & Electrical Engineering*, vol. 85, p. 106673, 2020. Article (CrossRef Link)

[5] J. C. Mukherjee and A. Gupta, "Distributed charge scheduling of plug-in electric vehicles using inter-aggregator collaboration," *IEEE Transactions on Smart Grid*, vol. 8, no. 1, pp. 331-341, 2017. Article (CrossRef Link)

[6] Y. Cao, T. Jiang, O. Kaiwartya, H. Sun, H. Zhou, and R. Wang, "Toward pre-empted EV charging recommendation through V2V-based reservation system," *IEEE transactions on systems, man, and cybernetics: systems*, vol. 51, no. 5, pp. 3026-3039, 2021. Article (CrossRef Link)

[7] S. Hou, C. Jiang, Y. Yang, and W. Xiao, "Electric Vehicle Charging Scheduling Strategy based on Genetic Algorithm," *Physics: Conference Series*, vol. 1693, no. 1, p. 012104, 2020. Article (CrossRef Link)

[8] C. Wang, C. Guo, and X. Zuo, "Solving multi-depot electric vehicle scheduling problem by column generation and genetic algorithm," *Applied Soft Computing*, vol. 112, p. 107774, 2021. Article (CrossRef Link)

[9] N. T. Milas, D. A. Mourtzis, P. I. Giotakos, and E. C. Tatakis, "Two-Layer Genetic Algorithm for the Charge Scheduling of Electric Vehicles," in *Proc. of 2020 22nd European Conference on Power Electronics and Applications (EPE'20 ECCE Europe)*, pp. P.1-P.10, 2020. Article (CrossRef Link)

[10] W.-J. Yin and Z.-F. Ming, "Electric vehicle charging and discharging scheduling strategy based on local search and competitive learning particle swarm optimization algorithm," *Journal of Energy Storage*, vol. 42, p. 102966, 2021. Article (CrossRef Link)

[11] X. Bai, Z. Wang, L. Zou, H. Liu, Q. Sun, and F. E. Alsaadi, "Electric vehicle charging station planning with dynamic prediction of elastic charging demand: A hybrid particle swarm optimization algorithm," *Complex & Intelligent Systems*, pp. 1035-1046, 2022. Article (CrossRef Link)

[12] N. Wang, B. Li, Y. Duan, and S. Jia, "A multi-energy scheduling strategy for orderly charging and discharging of electric vehicles based on multi-objective particle swarm optimization," *Sustainable Energy Technologies and Assessments*, vol. 44, p. 101037, 2021. Article (CrossRef Link)

[13] J. Garcia Alvarez, M. Á. González, C. Rodriguez Vela, and R. Varela, "Electric vehicle charging scheduling by an enhanced artificial bee colony algorithm," *Energies*, vol. 11, no. 10, p. 2752, 2018. Article (CrossRef Link)

[14] A. Chiş, J. Lundén, and V. Koivunen, "Reinforcement Learning-Based Plug-in Electric Vehicle Charging With Forecasted Price," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 3674-3684, 2017. Article (CrossRef Link)

[15] V. Moghaddam, A. Yazdani, H. Wang, D. Parlevliet, and F. Shahnia, "An online reinforcement learning approach for dynamic pricing of electric vehicle charging stations," *IEEE Access*, vol. 8, pp. 130305-130313, 2020. Article (CrossRef Link)

[16] N. Sadeghianpourhamami, J. Deleu, and C. Develder, "Definition and evaluation of model-free coordination of electrical vehicle charging with reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 203-214, 2020. Article (CrossRef Link)

[17] A. Marinescu, I. Dusparic, and S. Clarke, "Prediction-based multi-agent reinforcement learning in inherently non-stationary environments," *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, vol. 12, no. 2, pp. 1-23, 2017. [Article (CrossRef Link)](#)

[18] H. M. Abdullah, A. Gastli, and L. Ben-Brahim, "Reinforcement learning based EV charging management systems–a review," *IEEE Access*, vol. 9, pp. 41506-41531, 2021. [Article (CrossRef Link)](#)

[19] S. Dimitrov and R. Lguensat, "Reinforcement learning based algorithm for the maximization of EV charging station revenue," in *Proc. of 2014 International Conference on Mathematics and Computers in Sciences and in Industry*, pp. 235-239, 2014. [Article (CrossRef Link)](#)

[20] C. Jiang, Z. Jing, X. Cui, T. Ji, and Q. Wu, "Multiple agents and reinforcement learning for modelling charging loads of electric taxis," *Applied Energy*, vol. 222, pp. 158-168, 2018. [Article (CrossRef Link)](#)

[21] T. Qian, C. Shao, X. Wang, and M. Shahidehpour, "Deep reinforcement learning for EV charging navigation by coordinating smart grid and intelligent transportation system," *IEEE transactions on smart grid*, vol. 11, no. 2, pp. 1714-1723, 2020. [Article (CrossRef Link)](#)

[22] N. Mhaisen, N. Fetais, and A. Massoud, "Real-time scheduling for electric vehicles charging/discharging using reinforcement learning," in *Proc. of 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, pp. 1-6, 2020. [Article (CrossRef Link)](#)

[23] A. Chiş, J. Lundén, and V. Koivunen, "Scheduling of plug-in electric vehicle battery charging with price prediction," in *Proc. of Ieee pes isgt europe 2013*, pp. 1-5, 2013. [Article (CrossRef Link)](#)

[24] Z. Wen, D. O'Neill, and H. Maei, "Optimal demand response using device-based reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2312-2324, 2015. [Article (CrossRef Link)](#)

[25] H. Li, Z. Wan, and H. He, "Constrained EV charging scheduling based on safe deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2427-2439, 2020. [Article (CrossRef Link)](#)

[26] F. Zhang, Q. Yang, and D. An, "CDDPG: A deep-reinforcement-learning-based approach for electric vehicle charging control," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3075-3087, 2021. [Article (CrossRef Link)](#)

TIANYANG LI received the Ph.D. degree from the School of Computer Science and Technology, Harbin Institute of Technology, in 2018. He is currently a Lecturer with the School of Computer Science, Northeast Electric Power University. His research interests include service computing, cloud manufacturing, the Internet of Things, and intelligent optimization algorithms. E-mail: 872888861@qq.com

YINGNAN HAN received the bachelor's degree from the School of Computer Science and Technology, Changchun University, in 2020. He is currently working toward the master's degree in School of Computer Science, Northeast Electric Power University. His research interests include service computing, deep reinforcement learning, and intelligent optimization algorithms. E-mail: 2061331823@qq.com

XIAOLONG LI received Master's degree from the School of Economics and Management, Northeast Petroleum University in 2020. Currently, he is working at Harbin Branch of Bank of Inner Monglia as a retail customer manager. His research interests include service computing, big data applications, intelligent optimization algorithms, etc. E-mail: long1378568805@163.com