# A Data-Driven Approach for Real-Time Residential EV Charging Management

Zhiqiang Wan*, Hepeng Li†, Haibo He*, and Danil Prokhorov‡
*Department of Electrical, Computer and Biomedical Engineering, University of Rhode Island
RI 02881 USA, Email: {zwan, he}@ele.uri.edu
†Lab. of Networked Control Systems, Shenyang Institute of Automation, Chinese Academy of Sciences
Shenyang 110016 China, Email: cn.h.li@ieee.org
‡Toyota Research Institute North America
Ann Arbor, MI 48105 USA, Email: dvprokhorov@gmail.com

***Abstract*—When electric vehicle (EV) participates in demand response with real-time electricity price, the EV charging cost can be significantly reduced by properly managing the charging schedules according to these pricing data. However, due to the existence of randomness in the pricing process of the utility and the user's commuting behavior, determining a cost-efficient charging strategy becomes challenging. Traditional model-based solutions need a model to predict the uncertainty. Constructing a model-based controller is difficult when the heterogeneity of EV users is taken into consideration. In this paper, the EV charging management problem is formulated as an Markov Decision Process (MDP) which has unknown transition probability. A data-driven approach based on deep reinforcement learning is developed to determine the optimal EV charging strategy. The proposed approach does not need any system model information. Experimental results verify the effectiveness of our proposed approach.**

***Index Terms*—Data-driven, reinforcement learning, EV charging management.**

## I. Introduction

In recent years, EVs are becoming more and more popular because it can reduce the demand for fossil fuels. Since the real-time electricity price has been used by many utility companies [1], we can reduce the charging cost by properly managing the EV charging. However, due to the existence of randomness in user's commuting behavior and pricing process of the utility, efficiently managing the EV charging to reduce the charging cost becomes challenging.

Recently, numerous approaches in the literature have been developed to solve this problem. For instance, N. G. Paterakis et al. [2] considered the EV as a demand response resource and formulated its charging management as a convex optimization problem aiming to minimize the daily total operational cost. Yi Guo, et al. [3] developed a two-stage architecture for the cost-efficient operation of a microgrid-like EV parking deck which considers the intermittency of renewable outputs. S. I. Vagropoulos et al. [4] take the randomness in the regulation signals into consideration and implemented stochastic optimization to solve the EV charging management problem in both auxiliary service and day-ahead markets from the EV aggregator's perspective. Giulio Binetti et al. [5] proposed a formulation for the coordinated EV charging problem. In this formulation, the EV plug-in and plug-off frequency is considered, and a real-time greedy algorithm is developed to solve the this problem in a decentralized way. Y. T. Liao et al. [6] designed a stochastic dynamic programming approach that maximizes the profit of the operator with the real-time energy dispatch of an EV charging station which is equipped with photovoltaic panels. Qilong Huang et al. [7] proposed a Markov Decision Process (MDP) formulation for the EV charging management problem which takes the uncertainty and dynamics in wind energy supply into consideration. Then, a rollout algorithm is applied to achieve the optimal charging management policy. Although the above approaches achieve some success in EV charging management, they formulate this charging management problem as a model-based control problem that, in general, needs a model to predict the uncertainty. When constructing a model-based controller, it is difficult to choose a suitable model and estimate its parameters accurately. Moreover, if EV user's heterogeneity, such as different commuting behavior, are considered, different model parameters and even different models are needed.

In this paper, a data-driven approach based on deep reinforcement learning (RL) is proposed to overcome these disadvantages. Learning based approach does not need any information about the system model and has achieved great success in image processing [8], parameter estimation [9], [10], and complex decision-making applications, such as playing Atari 2600 [11] and Go game [12]. We formulate the EV charging management problem as an MDP from the user's perspective. Our aim is to determine a cost-efficient charging management policy which can take full advantage of the real-time electricity price. The proposed approach inputs the battery SOC and the past electricity prices, and outputs real-time EV charging schedules. Various experiments verify the effectiveness of our proposed approach.

The contributions of this paper are twofold. First, we formulate the EV charging management problem from the user's perspective by an MDP that has unknown transition probability. The randomness in the commuting behavior and the electricity price are considered. Second, a data-driven approach based on deep RL is developed to find an optimal policy to manage the EV charging.

978-1-5386-7703-2/18/$31.00 ©2018 IEEE

## II. Problem Formulation

In this paper, the EV charging management problem is formulated from the user's perspective. Our aim is to determine cost-efficient schedules that make full use of the real-time electricity price. The EV arrival time and departure time, and the daily energy consumption are supposed to change randomly because of the randomness in the traffic conditions and different commuting behavior.

A finite MDP with discrete time step $t = \{1, 2, \ldots, T\}$ is applied to formulate this problem as follow.

*1) State:* The system state $s_t$ is defined as $(u_t, E_t, P_{t-N}, \ldots, P_t)$. It contains three kinds of information: (1) $u_t$ denotes whether the EV is at home or not; (2) $E_t$ indicates the remaining energy of the EV battery; (3) $(P_{t-N}, \ldots, P_t)$ represents the previous N-hour's electricity prices.

*2) Action:* The action $a_t$ denotes the EV charging/discharging power. When the EV is charging, $a_t$ is positive. When EV is discharging energy back to the grid, $a_t$ is negative. In our formulation, there are three feasible actions, i.e., charging, discharging, and no action.

*3) State transition:* The state transition from the state $s_t$ to the next state $s_{t+1}$ is represented as

$$s_{t+1} = f(s_t, a_t, \omega_t), \tag{1}$$

where the state transition is both controlled by the action $a_t$ and influenced by the randomness $\omega_t$ related to the traffic conditions and commuting behavior.

*4) Reward:* At time step $t$, the reward is defined from user's perspective as

$$r_t = \begin{cases} -P_t * a_t, & t \neq t_\beta \\ -P_t * a_t - \tau * \left(E_{max} - E_t\right)^2, & t = t_\beta \end{cases} \tag{2}$$

where $t_\beta$ denotes the time step when the EV leaves home, $E_{max}$ represents the battery capacity, and $\tau$ is a coefficient. In this reward, $P_t * a_t$ indicates the charging cost, and $\tau * (E_{max} - E_t)^2$ is a penalty term which has a large value when the EV is not fully charged.

*5) Action-value function:* An action-value function $Q_{\boldsymbol{\pi}}(s, a)$ is proposed to measure the quality of a charging/discharging action $a$ under a given system state $s$. It is defined as the expected total sum of future rewards for the horizon of $K$ time steps as follow

$$Q_{\boldsymbol{\pi}}(s, a) = \mathbb{E}_{\boldsymbol{\pi}} \left[ \sum_{k=0}^{K} \gamma^k \cdot r_{t+k} \middle| s_t = s, a_t = a \right], \tag{3}$$

where $\boldsymbol{\pi}$ is the EV charging/discharging policy which maps from a system state $s$ to a charging/discharging action $a$, and the discount factor $\gamma$ is proposed to balance the trade-off between the future rewards and the immediate reward.

The aim of this charging management problem is to determine an optimal policy $\boldsymbol{\pi}^*$ to maximize the total rewards as

$$Q^*(s, a) = \max_{\boldsymbol{\pi}} Q_{\boldsymbol{\pi}}(s, a), \tag{4}$$

where $Q^*(s, a)$ denotes the optimal action-value function.

## III. Proposed Approach

Analytically determining the optimal policy $\boldsymbol{\pi}^*$ is challenging because the commuting behavior and the future electricity prices are not known in advance. In this paper, a reinforcement learning (RL) approach is proposed to solve this problem according to the Bellman equation [13]

$$Q_{i+1}(s, a) = \mathbb{E} \left[ r_t + \gamma \max_{a_{t+1}} Q_i(s_{t+1}, a_{t+1}) \middle| s_t = s, a_t = a \right]. \tag{5}$$

$Q(s, a)$ will converge to $Q^*(s, a)$, when the number of iterations $i \to \infty$ [14].

In this paper, a Fully-Connected Neural Network is proposed to estimate $Q^*(s, a)$, and this approach is called deep reinforcement learning (RL).

The overall architecture of the proposed approach based on deep RL is demonstrated in Fig. 1. The fully-connected neural network is applied to estimate the action-value of all feasible actions based on the electricity price and the battery SOC. The action with the largest action-value is outputted as the EV charging/discharging schedule.

### A. Architecture of Fully-Connected Neural Network

The fully-connected neural network has three-layers. This type of neural network can uniformly approximate any continuous function [15]. The input of the fully-connected neural network is the past 24-h electricity prices and the battery SOC. The input layer is connected to a hidden layer, and the value of the hidden unit can be calculated as

$$v = g\left(W_1 * x + b_1\right), \tag{6}$$

where $x$ denotes the past 24-h electricity prices and the battery SOC, $g$ represents the rectified linear activation function, and $W_1$ and $b_1$ are the parameters. Then, the hidden layer is connected to the output layer.

The output of the fully-connected neural network is the action-value of the feasible charging/discharging actions under system state $s$, i.e.,

$$Q(s, a) = g\left(W_2 * v + b_2\right), \tag{7}$$

where $W_2$ and $b_2$ represent the parameters. Then, the action with the largest action-value is outputted as the EV charging/discharging schedule.

### B. Training Process

Before deploying the proposed approach, we need to train the fully-connected neural network such that it can approximate the optimal action-value function $Q^*(s, a)$. The training algorithm is presented in Algorithm 1. At each time step, the charging/discharging action $a_t$ is selected according to the $\varepsilon$-greedy policy, i.e., the action is randomly selected with

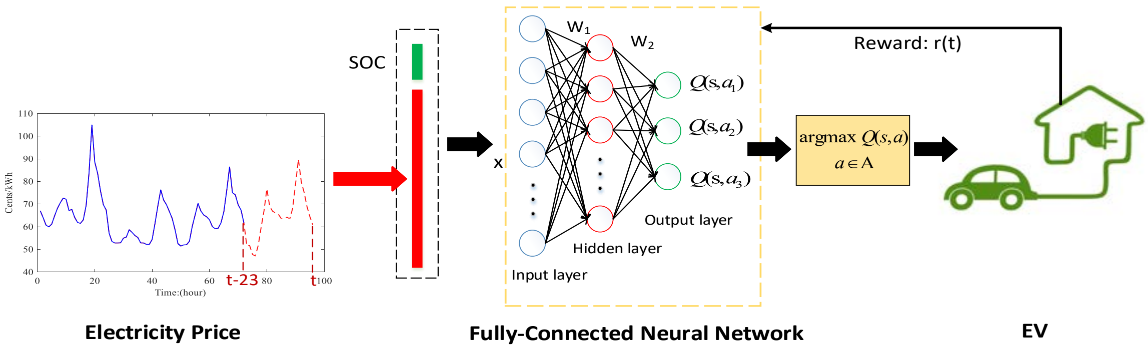


Fig. 1. The overall architecture of the proposed data-driven approach for EV charging/discharging management. The input of the fully-connected neural network is the past 24-h electricity prices and the battery SOC. Its output is the action-value for all feasible actions. The one with the largest action-value is outputted as the EV charging/discharging schedule.

**Algorithm 1** Training Process

Randomly initialize parameters $\theta$.
**for** Epoch=1:M **do**
  Initialize state $s_1$.
  **for** Time step t=1:T **do**
    Choose action $a_t$ based on $\varepsilon$-greedy.
    Take the selected action. Then, calculate reward $r_t$ and transit to the next state $s_{t+1}$.
    Store transition $(s_t, a_t, r_t, s_{t+1})$ in $\mathcal{D}$.
    Randomly sample minibatch of transitions $\mathcal{F} = \{(s_j, a_j, r_j, s_{j+1})\}_{j=1}^{\#\mathcal{F}}$ from $\mathcal{D}$.
    Calculate $y_j = \mathbb{E}_{s_{j+1}}\left[r_j + \gamma \max_{a'} Q\left(s_{j+1}, a'; \theta\right)\right]$.
    Optimize parameters $\theta$ by minimizing the loss function $\sum_{j=1}^{\#\mathcal{F}} \mathbb{E}\left[\left(y_j - Q\left(s_j, a_j; \theta\right)\right)^2\right]$.
  **end for**
**end for**

probability $\varepsilon$, otherwise the proposed approach is applied to choose the action. After taking the selected action, the reward $r_t$ is calculated, and the system is transited to state $s_{t+1}$. Then, $(s_t, a_t, r_t, s_{t+1})$ is stored in a replay memory $\mathcal{D}$. The parameters of the fully-connected neural network are optimized according to this replay memory. Specifically, a minibatch of transitions $\mathcal{F} = \{(s_j, a_j, r_j, s_{j+1})\}_{j=1}^{\#\mathcal{F}}$ are randomly sampled from the replay memory $\mathcal{D}$. Then, we can derive the loss function

$$L(\theta) = \sum_{j=1}^{\#\mathcal{F}} \mathbb{E}\left[\left(y_j - Q\left(s_j, a_j; \theta\right)\right)^2\right], \tag{8}$$

where $Q(s_j, a_j; \theta)$ denotes the action-value approximated by the fully-connected neural network, and $y_j = \mathbb{E}_{s_{j+1}}\left[r_j + \gamma \max_{a'} Q\left(s_{j+1}, a'; \theta\right)\right]$ represents the target action-value. Then, the parameters of the fully-connected neural network are optimized by minimizing the loss function $L(\theta)$.

**Algorithm 2** EV Charging/Discharging Managing

**Input:** Past 24-h electricity price and battery SOC.
**Output:** EV charging/discharging actions $a_{1:T}$.

**for** Time step t=1:T **do**
  Receive electricity price and battery SOC.
  Calculate action-value $Q(s_t, a; \theta)$.
  $a_t \longleftarrow \operatorname{argmax}_{a \in \mathcal{A}} Q(s_t, a; \theta)$
  Output EV charging/discharging action $a_t$.
**end for**

### C. Online Deployment

After the training phase, the parameters of the fully-connected neural network remain fixed during the online running phase. The proposed approach is deployed to generate real-time EV charging/discharging actions such that the charging cost is minimized. The online deployment is presented in Algorithm 2. Specifically, at time step $t$, the past 24-h electricity price and the battery SOC are fed into the fully-connected neural network to calculate the action-value for all feasible EV charging/discharging actions. Then, the action with the largest action-value is outputted for the EV charging/discharging management.

## IV. Experimental Results

In this section, the effectiveness of the proposed approach is demonstrated by simulation analysis. Section IV-A gives the information about the experimental setup. Section IV-B presents the details about the training process. Finally, in Section IV-C, the proposed approach is evaluated and compared with several benchmark methods.

### A. Experimental Setup

We evaluate the performance of our proposed approach in a real-world scenario. The hourly electricity price is downloaded from the California ISO [16]. It starts from February 1st, 2014 and lasts for 360 days. These days are divided into training and testing days. In each 30 consecutive days, the first 20 days are chosen as training data while the remaining 10 days are

TABLE I
RANDOM VARIABLES FOR COMMUTING BEHAVIOR.

| | Distribution | Boundary |
|---|---|---|
| Arrival time | $t_{arr} \sim \mathcal{N}\left(18, 1^2\right)$ | $15 \leq t_{arr} \leq 21$ |
| Departure time | $t_{dep} \sim \mathcal{N}\left(8, 1^2\right)$ | $6 \leq t_{dep} \leq 11$ |
| Battery SOC | $B_{SOC} \sim \mathcal{N}\left(0.5, 0.1^2\right)$ | $0.2 \leq B_{SOC} \leq 0.8$ |

used for testing. In each day, the time when the EV arrives and departs, and the remaining battery SOC when the EV arrives home are defined as random variables that obey truncated normal distributions. Their distributions are shown in Table. I. For the arrival time $t_{arr}$, its distribution $\mathcal{N}\left(18, 1^2\right)$ is bounded between 15 and 21. The departure time $t_{dep}$ is sampled from $\mathcal{N}\left(8, 1^2\right)$ and is bounded between 6 and 11. The battery SOC $B_{SOC}$ is sampled from $\mathcal{N}\left(0.5, 0.1^2\right)$ and is bounded between 0.2 and 0.8. In our experiments, we use a Nissan Leaf whose battery capacity $E_{max} = 24kWh$. When the EV arrives home, the charger has 3 charging/discharging levels (6 kW, 0 kW, -6 kW). The positive values refer to EV charging, and the negative values represent the discharging process.

The EV charging process is managed by our proposed deep RL based approach. The discount factor $\gamma = 1.0$ such that the immediate reward is as important as the future rewards. The coefficient $\tau$ is set to 12. In the fully-connected neural network, the number of hidden units is 64, and the number of output units is 3. The parameters of the fully-connected neural network are initialized from normal distributions and optimized by gradient descent in the training phase. The batch-size of transitions $\mathcal{F}$ sampled from the replay memory $\mathcal{D}$ for training is 32. A workstation with one NVIDIA TITAN Xp is applied for the training process. The proposed approach is implemented in Python with TensorFlow, a deep neural network package developed by Google Brain.

## B. Training Process

Our proposed deep RL based approach is trained for 40,000 epochs to learn the optimal EV charging/discharging management. The epoch starts at the time that the EV arrives home and ends when the EV leaves home. The cumulative rewards over 40,000 epochs is presented in Fig. 2. In the first 4,000 epochs, the charging/discharging action is randomly chosen from the 3 feasible actions. The replay memory $\mathcal{D}$ is filled up during this phase. The fully-connected neural network's parameters are not optimized in this phase. Then, from epoch 4,000 to epoch 8,000, the charging/discharging action is chosen based on the $\varepsilon-$ greedy policy. In this phase, the parameters start to update and $\varepsilon$ is decreased from 1.0 to 0.1, and keeps 0.1 afterward. In Fig. 2, we can observe that the cumulative rewards begin to increase gradually after epoch 8,000. Finally, the cumulative rewards converge at epoch 15,000. This result shows that our proposed deep RL based approach can learn a good policy to achieve high cumulative rewards.

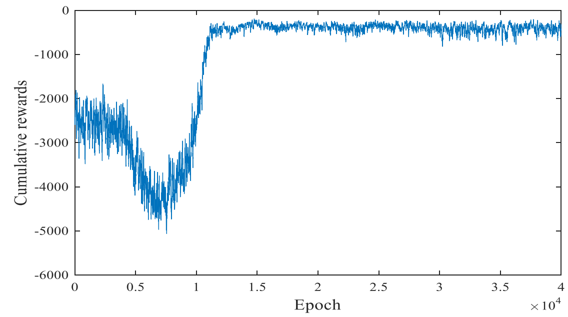


Fig. 2. The cumulative reward of each epoch during the training process.

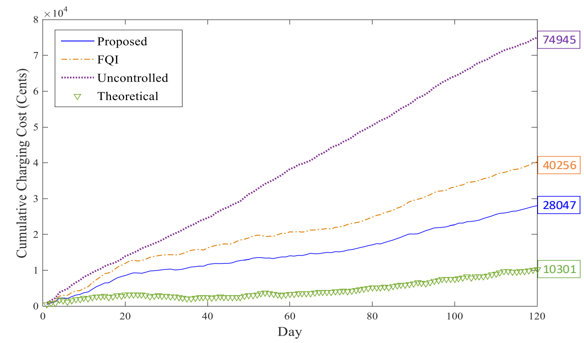


Fig. 3. Cumulative charging cost of the proposed and benchmark approaches over the 120 test days.

## C. Performance Evaluation

In order to verify the effectiveness of the proposed approach, we compare our deep RL based approach with several benchmark solutions, including fitted-Q iteration (FQI), uncontrolled strategy, and theoretical-limit strategy. For the theoretical-limit strategy, the time that the EV arrives and departs, the battery SOC, and the future 24-h electricity price are all available in advance. Therefore, the EV charging/discharging management problem is considered as a deterministic optimization problem. Then, the policy is determined by an optimization toolbox called YALMIP [17]. It is worth noting that this strategy gives a theoretical limit to the cumulative charging cost. However, it cannot be achieved in practice due to the randomness in the commuting behavior and the electricity price. For the FQI approach [18], the action-value function is approximated by a decision tree. Then, the action with largest action-value is outputted for the EV charging/discharging management. For the uncontrolled strategy, the EV is charged with the maximum charging rate once it arrives home.

In the performance evaluation, the evaluation metric is the cumulative charging cost over all the 120 test days. Fig. 3 shows the evaluation results of our proposed approach and the benchmark approaches. It can be observed that our proposed approach (blue solid line) greatly reduces the charging cost when compared to the uncontrolled strategy (purple dotted

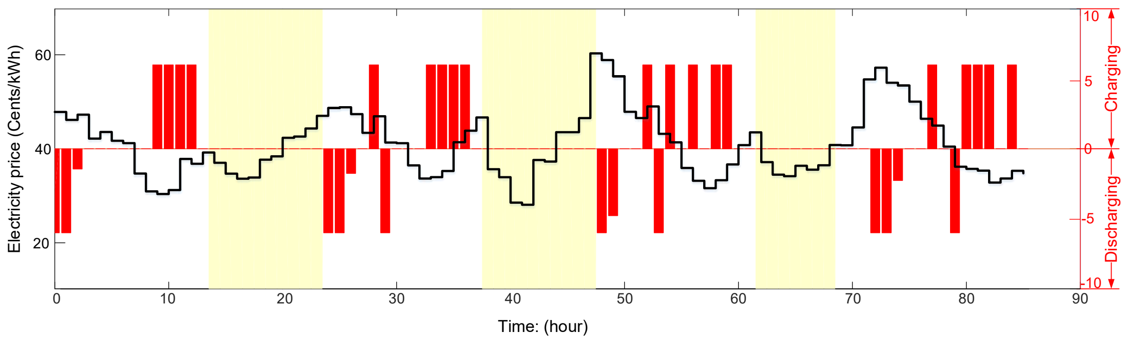


Fig. 4. Charging/discharging patterns of the proposed approach over 4 consecutive days. The yellow regions indicate that the EV leaves home.

line). Additionally, in comparison with FQI (orange dash-dotted line), our proposed approach is closer to the theoretical-limit strategy (green triangles). These results verify the effectiveness of our proposed approach.

In Fig. 4, the charging/discharging patterns over 4 consecutive days are presented to further investigate the performance of our proposed approach. The yellow regions shows the periods of time when the EV leaves home. The hourly electricity price (Cents/kWh) is represented by black line, and the charging/discharging energy (kWh) at each hour is indicated by the red bar. It can be observed that when the electricity price is low, our proposed approach learns to charge the EV. However, when the electricity price is high, the EV is discharged to earn revenue. These results verify that our proposed approach can efficiently manage the EV charging/discharging schedules according to the real-time electricity price such that the charging cost is reduced.

## V. Conclusion

In this paper, the residential EV charging/discharging management problem is formulated as an MDP which has unknown transition probability. The randomness of the commuting behavior and the electricity price are taken into consideration. A data-driven approach based on deep reinforcement learning has been developed to solve this problem. Our proposed approach does not require any system model information. In our proposed approach, a fully-connected neural network is applied to estimate the action-value of all feasible actions. The action with the largest action-value is outputted for the EV charging/discharging management. Experimental results demonstrate that our proposed approach outperforms numerous benchmark approaches.

## Acknowledgment

This work was supported in part by National Science Foundation (NSF) under grant CCF 1439011.

## References


[1] T. Namerikawa, N. Okubo, R. Sato, Y. Okawa, and M. Ono, "Real-time pricing mechanism for electricity market with built-in incentive for participation," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2714–2724, Nov 2015.

[2] N. G. Paterakis, O. Erdin, A. G. Bakirtzis, and J. P. S. Catalo, "Optimal household appliances scheduling under day-ahead pricing and load-shaping demand response strategies," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 6, pp. 1509–1519, Dec 2015.

[3] Y. Guo, J. Xiong, S. Xu, and W. Su, "Two-stage economic operation of microgrid-like electric vehicle parking deck," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1703–1712, May 2016.

[4] S. I. Vagropoulos and A. G. Bakirtzis, "Optimal bidding strategy for electric vehicle aggregators in electricity markets," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4031–4041, Nov 2013.

[5] G. Binetti, A. Davoudi, D. Naso, B. Turchiano, and F. L. Lewis, "Scalable real-time electric vehicles charging with discrete charging rates," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2211–2220, Sept 2015.

[6] Y. T. Liao and C. N. Lu, "Dispatch of ev charging station energy resources for sustainable mobility," *IEEE Transactions on Transportation Electrification*, vol. 1, no. 1, pp. 86–93, June 2015.

[7] Q. Huang, Q. S. Jia, Z. Qiu, X. Guan, and G. Deconinck, "Matching ev charging load with uncertain wind: A simulation-based policy improvement approach," *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1425–1433, May 2015.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[9] Z. Wan, H. He, and B. Tang, "A generative model for sparse hyperparameter determination," *IEEE Transactions on Big Data*, vol. PP, no. 99, pp. 1–1, 2017.

[10] Z. Wan, X. Hu, H. He, and Y. Guo, "A learning based approach for social force model parameter estimation," in *2017 International Joint Conference on Neural Networks (IJCNN)*, May 2017, pp. 4058–4064.

[11] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[12] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot *et al.*, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[13] R. Bellman, "Dynamic programming," 1958.

[14] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press Cambridge, 1998, vol. 1, no. 1.

[15] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals, and Systems (MCSS)*, vol. 2, no. 4, pp. 303–314, 1989.

[16] C. ISO, "http://oasis.caiso.com/mrioasis/logon.do."

[17] J. Lofberg, "Yalmip : a toolbox for modeling and optimization in matlab," in *2004 IEEE International Conference on Robotics and Automation (IEEE Cat. No.04CH37508)*, Sept 2004, pp. 284–289.

[18] D. Ernst, P. Geurts, and L. Wehenkel, "Tree-based batch mode reinforcement learning," *Journal of Machine Learning Research*, vol. 6, no. Apr, pp. 503–556, 2005.