35th International Electric Vehicle Symposium and Exhibition (EVS35)
Oslo, Norway, June 11-15, 2022

A Data-Driven Approach for Online EV Charging Management Considering Travel Pattern Heterogeneity

Xu Hao [1], Yue Chen [2*]
[1]*School of Mechanical Engineering, University of Science and Technology Beijing, Beijing 100083 China. Email: xuhao_research@ustb.edu.cn*
[2]*Department of Mechanical and Automation Engineering, the Chinese University of Hong Kong, ShaTin, NT, Hong Kong SAR, China. Email: yuechen@mae.cuhk.edu.hk*

Executive Summary

With the development of vehicle-to-grid (V2G) technologies, the electric vehicles (EVs) are able to sell energy back to the grid for peak shaving, congestion control, etc. A fundamental problem is how to properly charge/discharge the EVs to minimize their costs while satisfying their travel needs. This paper studies the charging management problem of massive EVs considering real-time electricity prices. A data-driven method based on deep Q-Network reinforcement learning is proposed for online implementation. Actual travel pattern heterogeneity in a typical city, Shanghai, is taken as an empirical example to illustrate the obtained charging/discharging strategy and the cumulative costs. The result shows a significant advantage of the proposed method over the uncontrolled charging, i.e., the proposed method can earn as much as 766 CNY during 4 months for the given test data with pre-decided departure time.

Keywords: V2G (vehicle to grid), BEV (battery electric vehicle), charging, Deep Q-Network, Travel pattern heterogeneity

1 Introduction

Massive deployment of electric vehicles (EVs) combined with the integration of clean renewable sources has been regarded as an effective way to reduce carbon emission. The worldwide stock of EVs is expected to account for 31% of the global fleet by 2040. China has also set an ambitious goal of having new energy vehicles, including EVs, account for more than half of total new vehicle sales by 2035 [1]. However, the uncontrolled charging of such a huge number of EVs may pose a risk of overburdening the power system. The vehicle-to-grid (V2G) technology can be a possible solution: by encouraging the EVs to adjust their charging and discharging behaviors, the power peak load can be smoothed [2]. Moreover, V2G enables the EVs to react to the floating electricity prices flexibly to decrease their usage costs. Therefore, V2G has attracted more and more attention. How to charge/discharge EVs properly in order to minimize the usage costs while satisfying their travel needs has become an important topic.

Despite the fruitful related work, most of them adopted offline models [3, 4] that assume complete information of the future. However, the fluctuating electricity prices caused by volatile renewable generations as well as the unpredictable travel behaviors make it impossible to get accurate predictions. Therefore, online charging algorithms without prediction are in great need. Typical algorithms are based on model predictive control [5], analysis of offline optimum structure [6], Lyapunov optimization [7], etc. The above algorithms all rely on an optimization model to generate the control strategy. However, considering the heterogeneity of EVs, fine-grained modeling of each EV can be difficult. To overcome this difficulty, reinforcement learning (RL), which is recently widely applied in automated vehicles [8], traffic planning [9], charging management [10], is adopted in this paper. However, the existing RL-based charging management works rarely take into account real-world travel pattern heterogeneity, especially that in China, the largest EV market.

This paper proposes a data-driven method for EV charging management based on deep Q-Network RL. Our main improvement over the existing work is the consideration of real-world travel pattern from Shanghai, including the daily vehicle kilometers traveled and the departure and arrival times, which is more practical. Moreover, we reveal that with known departure time, the performance of the proposed algorithm in terms of usage cost can be greatly improved. We also discuss the impact of prices on the charging/discharging behaviors.

This paper is organized as follows: Section 2 introduces the problem setup and the RL methodology. Section 3 gives the results and discusses the performance of the proposed model and algorithm. Finally we conclude this work in Section 4 with some potential implications and future research directions.

2 Problem Setup and Methodology

2.1 Problem Formulation

The EV charging management is modeled as a finite Markov Decision Process (MDP) with discrete time step $t = \{1, 2, ..., T\}$. Here the MDP assumes that the conditional probability distribution of future states depends only on the present state. Our aim is to determine cost-efficient charging schedules with limited past electricity prices and vehicle energy information. The MDP is defined as follows:

- **State**: The system state $s_t = \{h_t, E_t, P_{t-N}, ..., P_t\}$ consists of 1) indicator h_t that shows whether the EV is at home or not; here we assume that the EVs only charge or discharge when at home; 2) the state-of-charge (SOC) of the EV battery E_t; 3) the electricity prices for the former N hours $\{P_{t-N}, ..., P_t\}$. Here, the h_t unfolds over time sequentially, which means the EV owners don't know their departure time in advance. We make a comparison with the case when EVs have prior knowledge about the departure time in Section 3. The initial EV battery SOC E_1 is derived according to the actual travel needs and travel pattern.
- **Action**: The action a_t is the charging/discharging power of an EV, which can be positive (charging), negative (discharging), or zero (no action).
- **State Transition**: The state transition is formulated as

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

  where the battery energy in s_{t+1} is calculated as

$$E_{t+1} = \begin{cases} \max\{0, E_t + a_t\}, & \text{if } a_t < 0 \\ E_t, & \text{if } a_t = 0 \\ \min\{E_{max}, E_t + a_t\}, & \text{if } a_t > 0 \end{cases} \tag{2}$$

- **Reward**: Denote

$$\hat{a}_t = \begin{cases} \min\{a_t, E_{max} - E_t\}, & \text{if } a_t \geq 0 \\ -\min\{-a_t, E_t\}, & \text{if } a_t < 0 \end{cases} \tag{3}$$

Then at each time step t, the reward of a certain EV is defined as

$$r_t = \begin{cases} -P_t \cdot \hat{a}_t \cdot h_t \ , & t \neq t_\eta \\ -P_t \cdot \hat{a}_t \cdot h_t - \tau \cdot (E_{max} - E_t)^2 \ , & t = t_\eta \end{cases} \tag{4}$$

where t_η is the time when the EV leaves home and $\tau \cdot (E_{max} - E_t)^2$ denotes a penalty if the EV departs when it is not fully charged, which can be quite large if the SOC is low. Note that the actual penalty coefficient τ should be a comprehensive factor considering commute distance, charging preference, and range anxiety, which can be heterogeneous and time-varying. Here, a fixed τ is used for simplicity due to the lack of a reliable penalty function. The cumulative reward is $\sum_{t'=t}^{T} \gamma^{t'-t} r_{t'}$, where γ is the discount factor varying from 0 to 1.

- **Optimal action-value function**: The charging/discharging strategy is denoted by π, which is a collection of time-sequenced actions a. Let $Q^*(s,a)$ denote the maximum expected reward achievable at state s after taking action a.

$$Q^*(s,a) = \max_{\pi} \mathbb{E}_\pi \left[\sum_{t'=t}^{T} \gamma^{t'-t} r_{t'} \middle| s_t = s, a_t = a, \pi \right] \tag{5}$$

An iterative process is applied and a function approximator $Q_i(s,a;\theta)$ is used to estimate the action-value function in each iteration, so that we have $Q_i(s,a;\theta) \to Q(s,a;\theta) \approx Q^*(s,a)$.

2.2 Methodology and Data Sources

2.2.1 Mathematical Formulation and Methodology

In this paper, a fully-connected neural network, the Deep Q-Network (DQN), is adopted to as the function approximator $Q(s,a;\theta)$. Its update follows the Bellman equation [11]:

$$Q_{i+1}(s,a;\theta) = \mathbb{E} \left[r_t + \gamma \max_{a_{t+1}} Q_i(s_{t+1}, a_{t+1}; \theta) \middle| s_t = s, a_t = a \right] \tag{6}$$

$Q_i(s,a;\theta)$ will converge to $Q(s,a;\theta) \approx Q^*(s,a)$ finally. Here, θ is the parameter of the neural network.

To derive the θ and then to estimate the $Q^*(s,a)$, a fully-connected three-layer neural network is proposed. The input of the fully-connected neural network is the past 24-h electricity prices, the EV battery SOC calculated from the charging power and the daily vehicle kilometers travelled. The input layer is connected to a hidden layer, and the value of the hidden unit is calculated as

$$v = g(W_1 * x + b_1) \tag{7}$$

where we adopt the rectified linear unit g as the activation function between different layers. x denotes the aforementioned input information and v is an intermediate variable. Then the hidden layer is connected to the output layer. The output layer gives the action-values $Q(s,a)$:

$$Q(s,a) = g(W_2 * v + b_2) \tag{8}$$

θ is a collection of the parameters W_1, b_1, W_2, b_2. The action with the largest $Q(s,a)$ will be selected, i.e., $a^* = \text{argmax}_a Q(s,a)$. The framework of our proposed methodology is shown in Figure 1. The inputs of the model are the electricity price profile, the daily travel pattern, and the EV battery information. The charging/discharging schedule is the output.

2.2.2 Data Sources and Travel Pattern Heterogeneity

Real-world data is used in this paper. To be specific, the EV travel pattern heterogeneity, characterizing by the vehicle daily mileage, arrival time, and departure time, is obtained from an actual travel pattern analysis in Shanghai based on EV data acquired from the National Monitoring and Management Centre

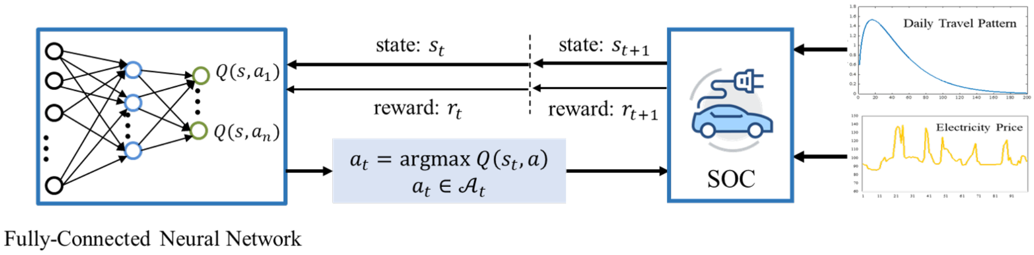

Figure 1: The framework of the proposed DQN method for EV charging management

for New Energy Vehicles (NMMC-NEV). The vehicle configurations, including the battery capacity and charging/discharging power, come from the China's electric vehicle market in 2021. In particular, the battery capacity $E_{max} = 49$ kWh, which is summarized from the loaded traction batteries data in 2021. When the EV arrives home, it has three possible actions $a_t = 7$ kW/0 kW/-7 kW, corresponding to different charging/discharging levels. Here, 7 kW is chosen since it is the common AC charging power in China regulated by the *Safety Requirements and Test Specifications of Electric Vehicle Conductive Supply Equipment* (GB/T 39752-2021). A positive a_t means that the EV is charging, a negative a_t means that the EV is discharging, and $a_t = 0$ means that it is neither charging nor discharging.

The initial EV battery SOC is decided by the vehicle daily mileage R. Usually, the actual travel pattern obeys the Gamma distribution [12] as verified by reference [13] using GPS-tracked driving data. This conclusion has been widely adopted in plug-in electric vehicle energy studies [14]. The Gamma distribution can be specified with two parameters: scale β and shape α:

$$f(r|\alpha, \beta) = \frac{\beta^{-\alpha} r^{\alpha-1} e^{-\frac{r}{\beta}}}{\Gamma(\alpha)} \tag{9}$$

The expectation and mode are two typical characteristic parameters to depict a distribution. The expectation of the Gamma distribution, which equals to $\alpha\beta$, can be estimated by dividing the annual driving distance by 365. The mode of the Gamma distribution is $(\alpha - 1)\beta$ and can be approximated by the daily round-trip distance to work or the most frequent destinations. The Gamma distribution we obtain from historical data is shown in Figure 2(a). We assume that the daily EV mileage is at least 50 km, otherwise, the EV owners may choose other means of transport, i.e., $R = \min\{f(r|\alpha, \beta), 50\}$. Then the initial EV battery SOC is calculated by

$$E_1 = E_{max} - 0.18 \times R \tag{10}$$

Here, 0.18 means the electricity consumption is 0.18 kWh/km. Besides, the departure and arrival times of EVs are randomly chosen from two truncated Normal distributions. The arrival time is fitted to a bounded Normal distribution $\mathcal{N}(18, 1^2)$ between 16 and 23 according to the commute time in Shanghai. Similarly, the departure time is fitted to the $\mathcal{N}(8, 1^2)$ and is bounded between 6 and 11 as shown in Figure 2(b).

Due to the lack of real-time electricity prices in China, the hourly Experiment Uniform Shanghai Energy Price (EUSEP) is derived from the Uniform Singapore Energy Price downloaded from the Energy Market Company, as shown in Table 1. The data covers 12 months, from January 1st, 2021 to December 31st, 2021. These days are divided into training and testing sets. The first 20 days in each month are set as training data and the rest days are adopted for testing. Note that the EUSEP increases a lot in 2021 due to the global energy crunch of high demand and tight supply conditions [15]. The electricity prices in Singapore is converted to those in Shanghai using a coefficient (0.502) derived from the ratio of average electricity price in Singapore and Shanghai, China [16]. Besides, the exchange rate is set as 6.34.

2.2.3 Training Process

The three-layer neutral network is trained to approximate the optimal action-value function $Q^*(s, a)$. To guarantee the balance between exploration and exploitation, ϵ-greedy policy is adopted, which selects a

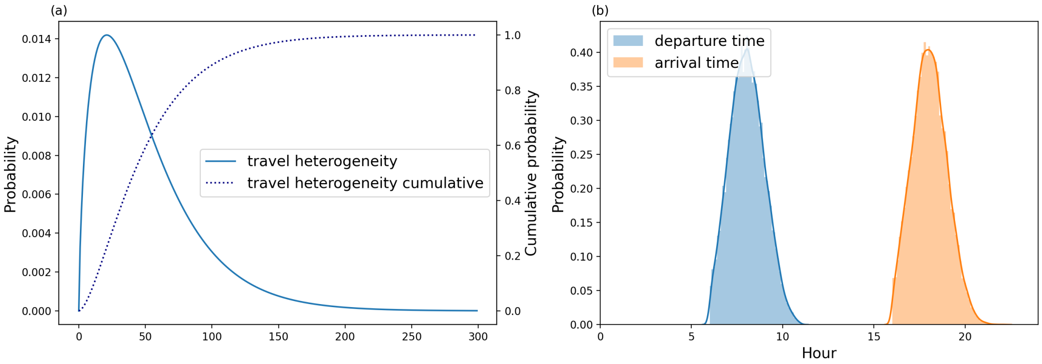

Figure 2: (a) Daily vehicle distance travelled distribution; (b) Travel time heterogeneity.

Table 1: Experiment Uniform Shanghai Energy Price in 2021

Period	Average EUSEP (CNY/kWh)
Jan-21	0.248
Feb-21	0.296
Mar-21	0.329
Apr-21	0.312
May-21	0.320
Jun-21	0.321
Jul-21	0.532
Aug-21	0.435
Sep-21	0.496
Oct-21	1.565
Nov-21	1.099
Dec-21	1.514

random action with the probability ϵ or otherwise selects the a_t that maximizes $Q(s_t, a_t; \theta)$ with probability $1 - \epsilon$. In this training, The ϵ begins at 1.0 and then decreases to and remains at 0.1 afterwards. At each step, the reward r_t is calculated from the chosen action, and the states moves to s_{t+1}. Then (s_t, r_t, a_t, s_{t+1}) is stored in a replay buffer. The loss function is:

$$L(\theta) = \mathbb{E}[(r + \gamma \max_{a'} Q(s', a'; \theta) - Q(s, a; \theta))^2] \tag{11}$$

To improve the computational efficiency, a mini-batch of transitions are randomly sampled from the replay buffer to update the parameters. The sizes of mini-batch and replay buffer are 64 and 2^{17}, respectively. The parameters are updated once per batch. Adam algorithm is used for parameter update.

Specifically, according to the effective time horizon function in [17], the discount factor γ is chosen as 0.9 since the EV owners only need to consider at most 24 steps in advance. The penalty coefficient τ is set to be 0.5 to avoid being caught in local optimum or taking a long time to converge. The learning rate is set as 0.001. The hidden units of the neural network are 64 to keep in line with the size of mini-batch, and the number of output units is 3. The input electricity prices are normalized by their 12-month means and standard deviation to speed up the convergence. The simulation is done with one MacBook with Apple M1 Max chip. The proposed approach is implemented in Python3.8 with TensorFlow, a deep neural network package developed by Google Brain.

After the training process, the proposed approach is deployed to generate real-time EV charging/discharging actions to minimize the usage cost as shown in Algorithm 1. It is assumed that EVs only charge at home and are plugged in once arrives home. For each epoch which starts at the time when the EV arrives home and ends when the EV leaves home, the inputs are the past 24-hour electricity prices and the initial battery SOC calculated from the electricity consumption and the daily vehicle kilometers travelled. Then the output is the action with the largest action-value calculated by the three-layer neural network. And then move to the next state.

Algorithm 1 EV Charging/Discharging Managing

Input: Past 24-hour electricity prices and initial battery SOC.
Output: EV charging/discharging actions $a_{t_1:t_\eta}$.
1: **for** $t = t_1$ to t_η **do**
2: Receive the electricity prices and initial battery SOC.
3: Calculate action-value $Q(s_t, a; \theta)$ from the neural network.
4: $a_t \leftarrow \text{argmax}_a Q(s_t, a; \theta)$.
5: $s_{t+1} = f(s_t, a_t)$.
6: **end for**

3 Results and Discussion

3.1 Results

Our proposed DQN approach is trained for 20,000 epochs. The cumulative rewards over 20,000 epochs is presented in Figure 3(a). As shown in the figure, the charging/discharging actions are randomly chosen in the first 2,000 epochs, and then the total rewards increase quickly and reach a relatively stable value with fluctuations. The trend is much more obvious in the average total rewards in Figure 3(b). It shows that a good policy with a high cumulative reward can be learned by the proposed method.

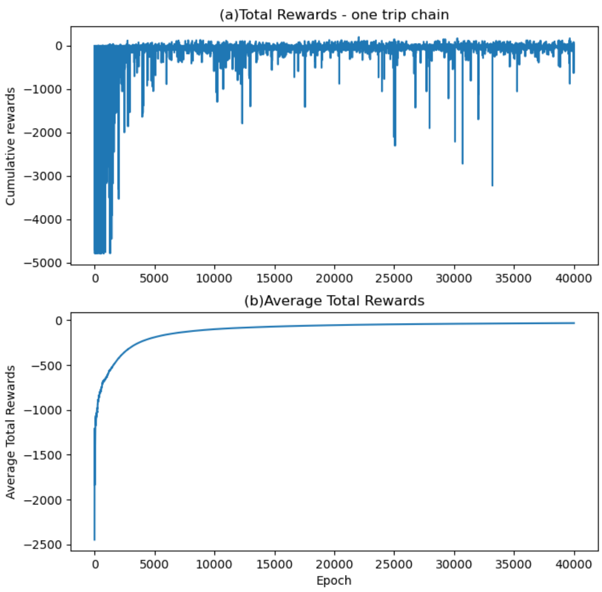

Figure 3: The cumulative reward of each epoch during the training process.

3.2 Performance Evaluation and Discussion

The proposed DQN-RL method is applied to the test data set and compared with the uncontrolled strategy, an improved DQN-RL method with known departure time, and the offline optimum. For the uncontrolled strategy, the EV is charged at the maximum charging rate until reaching its maximum battery capacity, i.e., 49 kWh in this research. For the improved DQN-RL method, the EV owners have a pre-decided departure time and the countdown hours is added to the expended state s_t. For the offline optimum, we assume complete information of the future, including the arrival and departure times of EVs, the initial battery SOC, and the future 24-hour electricity prices. A deterministic optimization is formulated to get the offline optimum, which is solved by the YALMIP toolbox in MATLAB. It is worth mentioning that the offline optimum can give the lowest cumulative cost and serve as a benchmark. However, the offline optimum is not practical due to the limited availability to the future information.

The cumulative costs of four different methods over 125 test days are shown in Figure 4. We can find that, compared with the uncontrolled strategy, the proposed DQN-RL method can greatly reduce the cost. The cumulative cost after 125 days under uncontrolled strategy is up to 1,256 CNY while that under DQN-RL is merely 18 CNY. Moreover, the improved DQN-RL has much lower cost than the original DQN-RL. This indicates that the effect of V2G can be largely improved if the EV owners can predict the departure time and take part in V2G accordingly. The cumulative cost of the improved DQN-RL method is -776 CNY, meaning that the EV owners can earn 776 CNY from the V2G. Offline optimum achieves the lowest cost. Due to the limited information, there is still an optimally gap between the offline optimum and the outcome of the DQN-RL method.

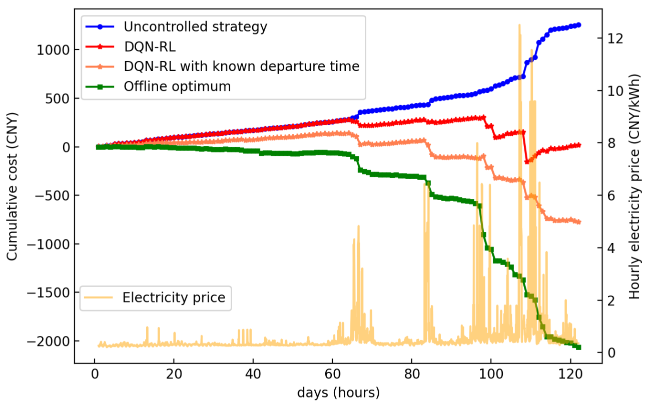

Figure 4: Cumulative charging costs over the 125 test days.

The charging/discharging profile generated by the DQN-RL method in three consecutive days is illustrated in Figure 5. The periods covered in light blue shadow are the time when the EV is not at home and so cannot be charged/discharged. The bars gives the charging/discharging energy (kWh) at each hour, and the red line is the change of EV battery SOC measured in the remaining energy. The proposed DQN-RL method can output an strategy that meets the travel needs derived from the real travel patterns.

4 Conclusion

In this paper, a Deep Q-Network based reinforcement learning method is proposed to solve the EV charging/discharging management problem in an online manner. The actual travel pattern heterogeneity in a typical city, Shanghai, is taken as an empirical example. The DQN-RL method is further improved by adding the known departure time to the states. The results reveal a significant advantage of the improved DQN-RL method over the uncontrolled charging method.

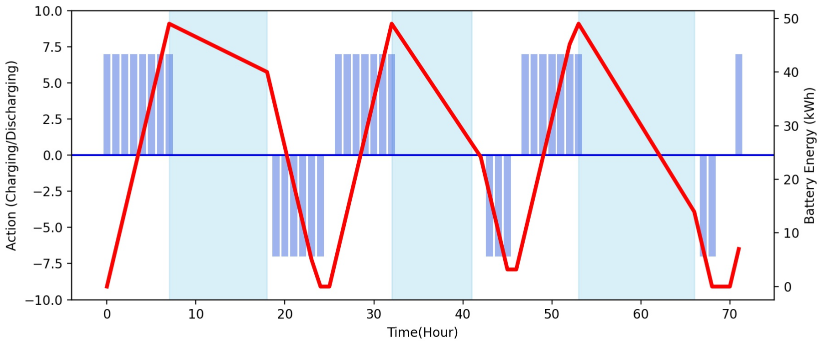

Figure 5: Charging/discharging actions profiles for 3 days

We have to admit that this paper still has a few limitations, such as the lack of actual electricity prices in China. The actual real-time electricity prices in China will be very helpful for us to accurately evaluate the economical effects of V2G in China. Moreover, the extra battery degradation is not taken into consideration in this paper, but in practice, high battery-swapping costs for the EV owners can occur if the V2G decreases the battery pack life.

In the future research, the action space can be changed from discrete space to continuous action space to reflect the variation of charging and discharging power. Moreover, this approach is the foundation to precisely evaluate the environmental and economical benefits of V2G.

Acknowledgments

Xu Hao's work was supported by the State Key Laboratory of Automotive Safety and Energy under Project No. KFY2220. Yue Chen's work was supported by CUHK Research Startup Fund. The authors are solely responsible for the views expressed in this study. Besides, the authors would like to thank Dr. Zhong Cao for his help in reinforcement learning knowledge.

References

[1] SAE China, *Energy-Saving and New Energy Vehicle Technology Roadmap*, 2nd ed. China Machine Press, 2021.

[2] N. Brinkel, W. Schram, T. AlSkaif, I. Lampropoulos, and W. Van Sark, "Should we reinforce the grid? cost and emission optimization of electric vehicle charging under different transformer limits," *Applied Energy*, vol. 276, p. 115285, 2020.

[3] Y. He, B. Venkatesh, and L. Guan, "Optimal scheduling for charging and discharging of electric vehicles," *IEEE transactions on smart grid*, vol. 3, no. 3, pp. 1095–1105, 2012.

[4] J. Wang, G. R. Bharati, S. Paudyal, O. Ceylan, B. P. Bhattarai, and K. S. Myers, "Coordinated electric vehicle charging with reactive power support to distribution grids," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 1, pp. 54–63, 2018.

[5] T. Li, B. Sun, Y. Chen, Z. Ye, S. H. Low, and A. Wierman, "Learning-based predictive control via real-time aggregate flexibility," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 4897–4913, 2021.

[6] L. Gan, U. Topcu, and S. H. Low, "Optimal decentralized protocol for electric vehicle charging," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 940–951, 2012.

[7] Y. Zhou, D. K. Yau, P. You, and P. Cheng, "Optimal-cost scheduling of electrical vehicle charging under uncertainty," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4547–4554, 2017.

[8] Z. Cao, D. Yang, S. Xu, H. Peng, and D. Zhao, "Highway exiting planner for automated vehicles using reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. PP, no. 99, pp. 1–11, 2020.

[9] D. Ma, B. Zhou, X. Song, and H. Dai, "A deep reinforcement learning approach to traffic signal control with temporal traffic pattern mining," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[10] Z. Wan, H. Li, H. He, and D. Prokhorov, "Model-free real-time ev charging scheduling based on deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5246–5257, 2018.

[11] R. Bellman, "Dynamic programming," *Princeton, USA: Princeton University Press*, vol. 1, no. 2, p. 3, 1957.

[12] X. Hao, H. Wang, Z. Lin, and M. Ouyang, "Seasonal effects on electric vehicle energy consumption and driving range: A case study on personal, taxi, and ridesharing vehicles," *Journal of Cleaner Production*, vol. 249, p. 119403, 2019.

[13] Z. Lin, D. Jing, C. Liu, and D. L. Greene, "Phev energy use estimation: Validating the gamma distribution for representing the random daily driving distance," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2287, no. 2287, 2012.

[14] X. Hao, Y. Yuan, H. Wang, and M. Ouyang, "Plug-in hybrid electric vehicle utility factor in china cities: Influencing factors, empirical research, and energy and environmental application," *eTransportation*, vol. 10, p. 100138, 2021.

[15] Energy Market Authority, "Electricity Prices on the Rise," Apr. 2022. [Online]. Available: https://www.ema.gov.sg/electricity-prices.aspx

[16] "Singapore electricity prices, September 2021," Sep. 2021. [Online]. Available: https://www.globalpetrolprices.com/Singapore/electricity_prices/

[17] J. Schulman, "The nuts and bolts of deep rl research," Technical report, Tech. Rep., 2016.

Presenter Biography

Xu HAO, is currently working as an assistant professor in School of Mechanical Engineering, University of Science and Technology Beijing. She received both Ph.D. and B.E. degrees in School of Vehicle and Mobility, Tsinghua University in China. Her research interest includes big data analysis of new energy vehicles, transportation decarbonization, market analysis of new energy vehicles.

Yue CHEN, is currently working as an assistant professor at the Department of Mechanical and Automation Engineering, the Chinese University of Hong Kong. She received both Ph.D. and B.E. degrees from the Department of Electrical Engineering, Tsinghua University, China. Her research interest includes smart grid, transactive energy, and integrated energy systems.