2023 IEEE 26th International Conference on
Intelligent Transportation Systems (ITSC)
24-28 September 2023. Bilbao, Bizkaia, Spain

Multi Actor-Critic PPO: A Novel Reinforcement Learning Method for Intelligent Task and Charging Scheduling in Electric Freight Vehicles Management*

Donghe Li[1,2], Chunlin Hu[1], Qingyu Yang[1,3], and Shitao Chen[4]

2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC) | 979-8-3503-9946-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ITSC57777.2023.10421834

***Abstract*— The rapid development of electric freight vehicles (EFVs) is driving the need for advanced management strategies, particularly given the dual demands of work scheduling and charging requirements. Amid this backdrop, the significance of intelligent scheduling algorithms has heightened, especially in the era of autonomous driving. In this study, we introduce a novel reinforcement learning (RL) strategy - the multi actor-critic proximal policy optimization (MAC-PPO) for the management of three categories of EFVs. Our approach utilizes distinct actor-critic networks for each category of EFVs, thereby creating a comprehensive and structured RL framework that effectively tailors task scheduling and charging strategies for different types of EFVs. Real-world conditions are emulated through the incorporation of a time-varying electricity price in our experiments. Results indicate that our methodology effectively optimizes the balance between freight tasks and charging demands. With increasing training episodes, we observe about 54%, 58%, and 60% reductions in average customer employment expenditure, average customer waiting time, and average charging expenditure, respectively. These findings underscore the efficiency and practicality of our proposed strategy in EFV management, reinforcing the pivotal role of intelligent scheduling in the autonomous driving age.**

I. Introduction

In recent decades, the field of intelligent transportation systems has experienced significant growth, driven by breakthroughs in autonomous driving, real-time communication, and positioning technologies [1], [2]. These advancements have revolutionized the way people and goods are transported, improving efficiency, safety, and convenience [3], [4]. Concurrently, the freight industry, a vital component of the global economy, is undergoing a major transformation. Traditional freight transportation methods are being replaced by more automated, eco-friendly, and efficient solutions, such as EFVs [5]. Supported by advances in battery technology and charging infrastructure, these vehicles are increasingly being adopted in logistics and transportation sectors, providing significant benefits such as reduced emissions, lower fuel costs, and decreased noise pollution [6], [7]. As autonomous driving technologies continue to advance, the potential for implementing unmanned freight services in the near future grows, promising to further enhance efficiency and sustainability [8].

To accommodate the widespread adoption of EFVs, effective scheduling and dispatch systems are essential. These systems must address two primary requirements: charging and freight pick-up. Like other electric vehicles, electric trucks require regular charging to ensure optimal performance [9]. Integrating charging into dispatch systems is crucial for maintaining seamless operations and minimizing downtime. Additionally, the dynamic nature of the freight industry demands flexible and efficient pick-up scheduling to meet diverse customer needs, highlighting the importance of automated dispatching in the context of electric freight vehicles [10].

However, developing efficient scheduling and dispatch systems for EFVs presents several unique challenges. First, unlike other vehicles, EFVs come in different types, each with its own charging needs and capacities. This variety necessitates careful planning and optimization to minimize delays and ensure timely deliveries [11]. The limited range and longer charging times of EFVs, compared to traditional vehicles, add to the complexity. Additionally, the availability and capacity of charging infrastructure can vary significantly across different regions, further complicating the scheduling process [12]. Second, the dynamic nature of freight demand, characterized by fluctuating pick-up requests, time-sensitive deliveries, and varying load sizes, presents a complex optimization problem. Furthermore, as EFVs become more prevalent, there will be increased competition for charging resources, necessitating intelligent algorithms capable of coordinating multiple vehicles and balancing the competing demands of charging and freight pick-up [13].

In light of these challenges, traditional optimization techniques often struggle to address the dynamic environment and complex demands associated with electric freight vehicle scheduling and dispatch. These methods typically rely on static models and assumptions, which may not accurately capture the continuously changing nature of freight demand and the operational constraints of electric vehicles. Consequently, there is a growing need for more adaptive and flexible approaches to tackle the unique challenges associated

*The work was supported in part by the National Science Foundation of China under Grants 61973247 and 62203350, in part by China Postdoctoral Science Foundation 2021M692566, in part by the Key Program of the National Natural Science Foundation of China under Grant 61833015, in part by the operation expenses for universities' basic scientific research of central authorities xzy012021027, in part by the Key Research and Development Program of Shaanxi under Grant 2022GY-033

[1]School of Automation Science and Engineering, Xi'an Jiaotong University, China. E-mails: lidonghe2020@xjtu.edu.cn, hucl0918@stu.xjtu.edu.cn

[2]MOE Key Laboratory for Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an, China

[3]SKLMSE lab, Xi'an Jiaotong University, Xi'an 710049, China. E-mail: yangqingyu@mail.xjtu.edu.cn

[4]College of Artificial Intelligence, Xi'an 710049, China. E-mail: chenshitao@xjtu.edu.cn

with EFV management. RL has emerged as a promising solution for addressing these challenges, as it is well-suited to handle complex optimization problems in dynamic environments [14], [15]. However, typical RL approaches may not be directly applicable to the electric freight vehicle scheduling problem due to the heterogeneous nature of the actions involved, such as charging and picking up freight. These diverse actions often require specialized handling and adaptation, as conventional RL methods might struggle to effectively manage the intricacies of such a problem [16]. Therefore, developing tailored RL algorithms that can effectively account for the specific challenges associated with electric freight vehicle management is crucial for achieving efficient scheduling and dispatch in this context.

To address the above two challenges, we developed the MAC-PPO EFV management method, a novel approach that incorporates RL into EFV task and charging scheduling. The contributions can be summarized as the following three aspects:

- We've created a comprehensive EFV task and charging optimization model that captures the unique aspects of EFV operations, including types of vehicles, charging infrastructure, and freight demand dynamics.
- We've presented a MAC-PPO EFV management method for the above optimization model. The proposed unique multi actor-critic architecture allows for optimized actions for different types of EFVs, thus improving overall system efficiency.
- We evaluate our proposed approach by using a real-world data. The results demonstrate its capability in managing dynamic freight demands and varying charging constraints.

The remainder of this paper is organized as follows. The model and formalization are presented in Section II. The MAC-PPO EFV management method is proposed in Section III. And numerical experiments are performed to validate the effectiveness of the proposed method in Section IV. Finally, some concluding remarks are given in Section V.

II. Models and Formalization

In this section, we first introduce the system model of our paper, provide the notations and assumptions used in our scheduling model, and then detail the problem formalization.

A. System Model

As depicted in Fig. 1, three key components constitute the EFV scheduling system: the Scheduling Center (SC), EFVs, and freight (or order). Within this framework, the SC observes the battery status of all EFVs and customer requests, scheduling suitable EFVs to transport freight. EFVs, acting in response to SC's orders, submit their current battery status and charging requests to SC. Freight, acting as a consumer, provides its information to the SC and requests support. The proposed system model captures the interactions between the main components, providing a foundation for designing a RL approach to schedule EFVs.

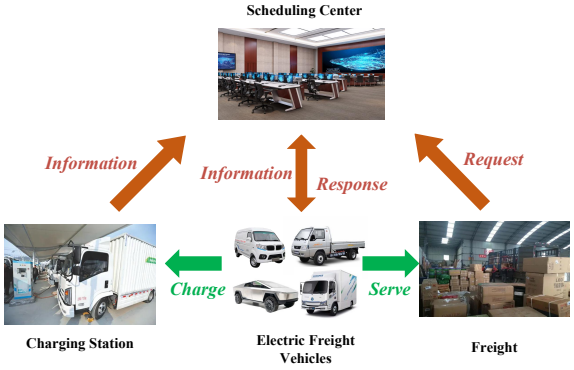


Fig. 1: System model

B. Notations and Assumptions

The time horizon is divided into T time slots. To utilize EFVs efficiently, we consider three types of EFVs in our paper: small EFVs, medium EFVs, large EFVs, represented by the set V. We assume that all EFVs travel at a fixed speed, represented by v. M denotes the mass of freight, and D denotes the distance between the starting and ending location of an order. The maximum load capacities of small, medium, and large EFVs are represented by m_1, m_2, and m_3, respectively, while their respective charging speeds are denoted by C_1, C_2, C_3.

C. Problem Formalization

We aim to optimize three objectives in this paper: reducing the expenditure of customers, minimizing customer waiting time, and decreasing EFVs' charging expenditure. The customer expenditure primarily refers to the cost of employing EFVs. Assuming a customer employs a_s small EFVs, a_m medium EFVs and a_l large EFVs, and all freight will be transported to the destination after k trips with these EFVs, the customer's expenditure can be formulated as:

$$Q_f^t = k(a_s Q_1^t(D) + a_m Q_2^t(D) + a_l Q_3^t(D)) \tag{1}$$

where $Q_1^t(D)$, $Q_2^t(D)$, $Q_3^t(D)$ denotes the employing cost of each small EFVs, medium EFVs and large EFVs in a travelling, respectively. $Q_i^t(D)$ could be expressed as follows:

$$Q_i^t(D) = \begin{cases} Q_i(0), & D \le D_i \\ Q_i(0) + d_i(D - D_i), & D > D_i \end{cases} \tag{2}$$

where $i \in \{1, 2, 3\}$, $Q_i(0)$ denotes starting expenditure, d_i denotes the expenditure each unit distance beyond starting distance.

The transportation times k must meet

$$\begin{cases} (k-1)(a_s m_1 + a_m m_2 + a_l m_3) < M \\ k(a_s m_1 + a_m m_2 + a_l m_3) \ge M \end{cases} \tag{3}$$

This paper does not consider the time that EFVs leave and return the charging station. Thus, the waiting time of the customers mainly refers to the time period between a group of EFVs firstly reach the starting location of an order and

finally leave the ending location, which could be formulated as follows:

$$T_w^t = \frac{(2k-1)D}{v} \tag{4}$$

Charging expenditure of EFVs refers to the cost that the EFVs charge in charging station. This paper assumes all EFVs, that do not participate in transportation assignment, charge in charging station. If there are c_s small EFVs, c_m medium EFVs and c_l large EFVs in the charging station at time step t, the total charging expenditure could be formulated as:

$$Q_c^t = c_s Q_1^t(C) + c_m Q_2^t(C) + c_l Q_3^t(C) \tag{5}$$

where $Q_1^t(C)$, $Q_2^t(C)$, $Q_3^t(C)$ denotes the charging cost of each small EFVs, medium EFVs and large EFVs, respectively. To stay close to reality, a time-varying electricity price is introduced. The charging cost of each EFV could be formulated as:

$$Q_j^t(C) = \eta^t C_j \tag{6}$$

where η^t indicates the electricity price at time slot t, $j \in \{1, 2, 3\}$.

Then, satisfactory of assignment (SOA) and satisfactory of charging (SOC) are introduced to describe our problem. SOA refers to the satisfactory of an EFV to customer, which could be formulated as:

$$SOA_k^t = \omega_1 Q_i^t(D) + \omega_2 T_w^t \tag{7}$$

where $i \in \{1, 2, 3\}$ and is decided by k. ω_1 and ω_2 are parameters.

SOC is influenced by electricity price, which could be formulated as:

$$SOC_k^t = \omega_3 Q_j^t(C) \tag{8}$$

where $j \in \{1, 2, 3\}$ and is decided by k. ω_3 is a parameter.

The final problem formalization is as follows:

$$\min_{a_{1k}^t, a_{2k}^t} \sum_{t=0}^{T} \sum_{k \in V} (a_{1k}^t SOA_k^t + a_{2k}^t SOC_k^t), i, j \in \{1, 2, 3\} \tag{9}$$

$$a_{1k}^t, a_{2k}^t = \{0, 1\}, \forall k \in V \tag{10}$$

$$a_{1k}^t + a_{2k}^t = 1, \forall k \in V \tag{11}$$

where a_{1k}^t and a_{2k}^t are optimization parameters, and represent the status of an EFV at time step t. k represent the ID of an EFV in V.

The objective function (Equation (9)) is composed of two terms representing the satisfaction of assignment and satisfaction of charging of each EFV, respectively. The first constraint (Equation (10)) denotes that the optimization parameter values are binary, where $a_{1k}^t = 1, a_{2k}^t = 0$ indicates an EFV is transporting freight and $a_{1k}^t = 0, a_{2k}^t = 1$ indicates an EFV is charging. The second constraint (Equation (11)) indicates that an EFV can only be in one state—either transporting or charging—at any given time.

III. Our Approach

In this section, we first give an overview of our proposed MAC-PPO EFV management method. Then, the Markov frame of proposed approach is introduced. Finally, the detail MAC-PPO is introduced.

A. Overview of the Multi Actor-Critic Proximal Policy Optimization EFV Management Method

The complexity and variability of the freight scheduling problem for different types of EFVs necessitate a dynamic and adaptable solution. To this end, we leverage the power of RL, which is adept at learning optimal policies from complex environments through interaction. Specifically, we propose a novel RL-based approach, the MAC-PPO EFV management method, which extends the traditional PPO algorithm by introducing three pairs of actor-critic networks. Each pair of networks is responsible for a specific type of vehicle, which greatly enhances the adaptability and performance of our approach in different scenarios.

Fig. 2 illustrates the structure of the proposed MAC-PPO EFV management method. The SC is modeled as an agent interacting with the environment. At each time step, the SC observes the environment state s_t, which is then fed into three pairs of old actor networks and actor networks. Each actor network generates an action and a probability for that action, while each old actor network produces a value for the current state-action pair. The action determines the number of small, medium, and large EFVs needed for the order. After the action is executed, the agent transitions to the next state s_{t+1} and receives a reward r_t. The state, action, probability, value, and reward are stored in a buffer and used to update the parameters of the actor networks with the output of the three critic networks. This structure significantly boosts the decision-making capability of our approach by enabling it to consider different types of vehicles independently and simultaneously.

B. Markov Decision Process Formulation

1) State: The environment state is designed as $s_t = \{t, N_s, N_m, N_l, M, D\}$, where t denotes the current time slot and $t \in \{0, 1, \cdots, T\}$. N_S, N_m, N_l denotes the number of small EFVs, medium EFVs and large EFVs in the charging station at time step t, respectively. M, D denotes respectively the mass and travelling distance of the order.

2) Action: In this paper, a_t is a set with three dimensions, which is formulated as:

$$a_t = \left\{a_t^1, a_t^2, a_t^3\right\} \tag{12}$$

where $a_t^1 \in A_1$, $a_t^2 \in A_2$, $a_3^t \in A_3$ is the output of three actor networks, indicating the number of three types of EFVs arranged to transport freight by SC. A_1, A_2 and A_3 are respectively as follows:

$$A_1 = \{0, 1, \cdots, m\} \tag{13}$$

$$A_2 = \{0, 1, \cdots, n\} \tag{14}$$

$$A_3 = \{0, 1, \cdots, q\} \tag{15}$$

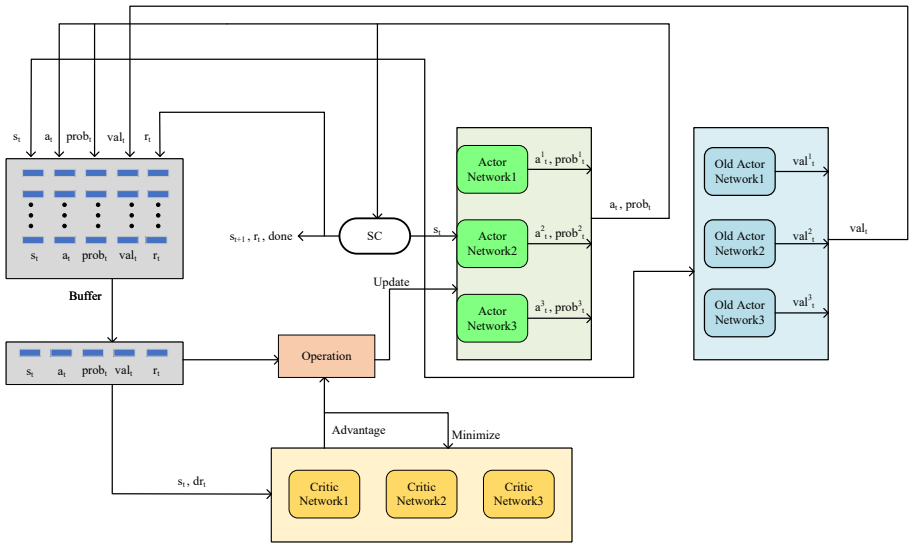


Fig. 2: Structure of MAC-PPO EFV Management method

Then, the number of EFVs leaving the charging station is respectively $a_s = a_t^1$, $a_m = a_t^2$ and $a_l = a_t^3$, and the number of EFVs in the charging station is respectively $c_s = N_s - a_t^1$, $c_m = N_m - a_t^2$ and $c_l = N_l - a_t^3$.

3) Reward: According to a_t, Q_f^t, T_w^t and Q_c^t could be obtained by using Equation (1), Equation (4), Equation (5). r_t denotes the reward that the agent gets after taking an action at time step t.

The reward in this paper is formulated as:

$$r_t = -(\omega_1 Q_f^t + \omega_2 T_w^t + \omega_3 Q_c^t) \tag{16}$$

Then, the cumulative reward of one day can be formulated as follows:

$$r = \sum_{t=0}^{T} r_t = -\sum_{t=0}^{T} (w_1 Q_f^t + w_2 T_w^t + w_3 Q_c^t) \tag{17}$$

The meaning of Equation (17) and Equation (9) is consistent. Thus, the optimization objective in this paper can be expressed as the maximum cumulative reward in one day.

C. Multi Actor-Critic Proximal Policy Optimization

Our approach adopts the PPO algorithm to train the three pairs of actor-critic networks. PPO, a policy gradient algorithm, incorporates an actor network and a critic network. In our MAC-PPO method, the actor network generates a policy that the agent follows to take actions, while the critic network estimates the value of each state, aiding in the policy update. The MAC-PPO distinguishes itself from the standard PPO by using three pairs of actor-critic networks, each dedicated to a specific type of vehicle.

In MAC-PPO, two policies are used: the new policy $\pi_\theta(a_t|s_t)$ and the old policy $\pi_{\theta_{old}}(a_t|s_t)$, where θ denotes the policy parameter vector. The probability ratio $r_t(\theta)$ is then expressed as:

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)} \tag{18}$$

One of the key features of PPO, and consequently of MAC-PPO, is the introduction of the advantage function $\hat{A}_t$, which is formulated as:

$$\hat{A}_t = \delta_t + (\gamma\lambda)\delta_{t+1} + \cdots + (\gamma\lambda)^{T-t+1}\delta_{T-1} \tag{19}$$

$$\delta t = Q(s_t) - V(s_t) = r_t + \gamma V(st+1) - V(s_t) \tag{20}$$

Unlike the trust region policy optimization algorithm that uses KL divergence to compute the policy gradient, PPO and MAC-PPO introduce a clip operation to calculate the policy gradient:

$$L_\pi^{clip}(\theta) = \hat{E}_t[\min(r_t(\theta)\hat{A}_t, clipr_t(\theta), 1-\varepsilon, 1+\varepsilon] \tag{21}$$

where ε is a hyperparameter used to clip the value of $r_t(\theta)$. This approach removes the incentive for $r_t(\theta)$, which is not in the interval $[1-\varepsilon, 1+\varepsilon]$.

The comprehensive Markov Decision Process (MDP) of our MAC-PPO approach is described in Algorithm 1. By integrating three pairs of actor-critic networks into the PPO framework, our MAC-PPO method effectively handles the multi-vehicle type problem and significantly improves the efficiency of the freight scheduling process.

IV. Performance Evaluation

This section introduces the evaluation settings and subsequently presents a series of experiments conducted to demonstrate the effectiveness of the approach proposed in this study.

A. Settings

In the simulation experiments, we apply a real data about EFVs from 2022. The number T of time slots is set to 24. According to our survey, a new order will appear approximately every one or two hours, and the SC only schedules EFVs at integer moment of time. Thus, the environment updates its state each a random time between one and two hours. The initial number of small EFVs, medium EFVs, large EFVs

Algorithm 1: MDP of our approach

Input: Episode number e, time slot number T for each episode

Load the parameters of trained actor networks;
for $episode = 0$ *to* e **do**
  Initialize environment state s_0;
  Initialize total reward r of each episode;
  for $t = 0$ *to* T **do**
    Obtain environment state s_t;
    Obtain an action a_t^1, a probability pro_t^1, a value val_t^1 ;
    Obtain an action a_t^2, a probability pro_t^2, a value val_t^2 ;
    Obtain an action a_t^3, a probability pro_t^3, a value val_t^3 ;
    Interact with environment and get a reward r_t;
    Update environment state and obtain s_{t+1};
    Store transition $\{s_t, a_t^1, prob_t^1, val_t^1, r_t\}$;
    Store transition $\{s_t, a_t^2, prob_t^2, val_t^2, r_t\}$;
    Store transition $\{s_t, a_t^3, prob_t^3, val_t^3, r_t\}$;
    $s_t = s_{t+1}$;
    $r = r + r_t$;
  end
  Update the actor network;
  Update the critic network;
end
return Scheduling result

are set to 10, 20 and 10, respectively. The maximum load capacity m_1, m_2, m_3 are set to 1t, 1.5t and 8t, respectively, and the charging speed C_1, C_2, C_3 are set to 80%, 70% and 50% every hour. D_1, D_2, D_3 are set to 5km, 5km and 15km, respectively, and $Q_1(0)$, $Q_2(0)$, $Q_3(0)$ are set to 55Yuan, 100Yuan and 280Yuan. M, D are both random integers in $(0, 100)$. ω_1, ω_2, ω_3 are respectively as 0.4, 0.3 and 0.3. The electricity price is shown in Fig. 3.

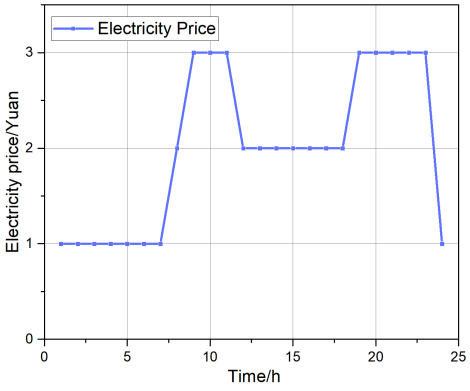


Fig. 3: Electricity price

The MAC-PPO algorithm is used to train the data, which includes three actor networks and three critic networks. Each network features two hidden layers. There are 7 neurons, 128 neurons, and 1 neuron in the input, hidden, and output layers of each network, respectively. The number of training episodes e is set to 1000. The learning rate and discount factor are respectively set to 0.01 and 0.98. The memory size and batch size are set to 10,000 and 64, respectively. During training, the environment state is updated every one or two hours, and the parameters of actor and critic networks are updated every 10 episodes.

B. Reinforcement Learning Training Performance

This section evaluates the performance of the approach proposed in this study. The performance factors include an overall evaluation, customer evaluation, and EFV evaluation.

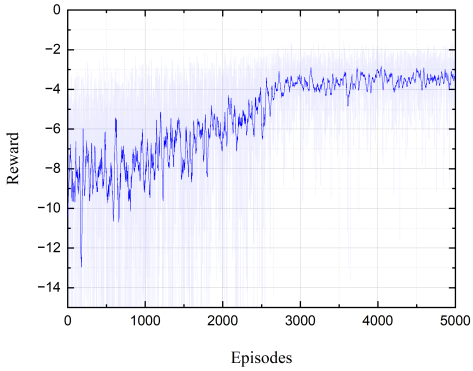
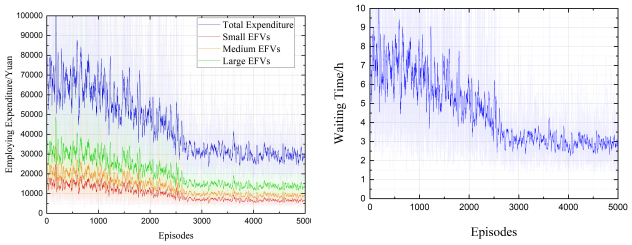


Fig. 4: Reward

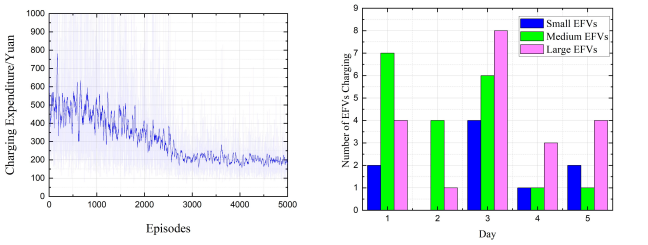
As illustrated in Fig. 4, the cumulative reward is relatively low in the initial training episodes. With more training episodes, the cumulative reward gradually increases and ultimately stabilizes at a high level. The upward trend in the graph demonstrates the effectiveness of our approach in optimizing the overall system performance.



(a) Customer expenditure (b) Customer waiting time

Fig. 5: Customer evaluation

1) Customer Evaluation: Fig. 5-(a) to (b) present the performance from the perspective of customers. As depicted in Fig. 5-(a), the total expenditure incurred by customers decreases by nearly 54% over the training episodes. The large EFVs have the highest expenditure, whereas the small EFVs incur the lowest expenditure. Fig. 5-(b) shows that the average waiting time for customers starts at approximately 7 hours per order during initial training and reduces to nearly 3 hours per order after convergence. Both figures illustrate that our proposed approach effectively considers customer satisfaction.



(a) EFV charging expenditure (b) Number of EFVs in the charging station

Fig. 6: EFV Evaluation

2) EFV Evaluation: Fig. 6-(a) illustrates the average daily charging expenditure for all EFVs. At the onset of training, the expenditure decreases from nearly 500 Yuan each day and eventually converges to 200 Yuan each day, marking a reduction of almost 60%. This demonstrates that our approach can effectively adjust the number of EFVs charging in the charging station according to the prevailing electricity prices. To further understand the scheduling status of all categories of EFVs, we analyzed the number of EFVs at the charging station at five random times. The results, presented in Fig. 6-(b), reveal that fewer than half of all types of EFVs with the lowest battery energy are charging at most times. This arrangement ensures that most EFVs are deployed to transport freight, thereby reducing customer waiting times.

In summary, our proposed approach reduces the average charging expenditure and optimally schedules EFVs to charge at appropriate times. It also considers the satisfaction of the EFV enterprise.

V. Conclusion

In conclusion, this work offers a significant contribution to the field of EFVs management by presenting a unique approach - the multi actor-critic proximal policy optimization. This method is a structured RL framework that deploys three distinct actor-critic networks for the effective management of three categories of EFVs. It optimally tailors task scheduling and charging strategies for different types of freight vehicles, creating a balance between freight tasks and charging demands. The effectiveness of our approach has been confirmed through simulation experiments, showing notable reductions in average customer employment expenditure, average customer waiting time, and average charging expenditure. These promising results underscore the value and practicality of our proposed method in EFV management, highlighting the potential of intelligent scheduling strategies in the era of autonomous driving. Future work could explore the application of our method in different contexts and with various types of electric vehicles to further validate its versatility and scalability.

References

[1] E. Wang, R. Ding, Z. Yang, H. Jin, C. Miao, L. Su, F. Zhang, C. Qiao, and X. Wang, "Joint charging and relocation recommendation for e-taxi drivers via multi-agent mean field hierarchical reinforcement learning," *IEEE Transactions on Mobile Computing*, vol. 21, no. 4, pp. 1274–1290, 2020.

[2] Y. J. Jang and Y. D. Ko, "System architecture and mathematical model of public transportation system utilizing wireless charging electric vehicles," in *2012 15th International IEEE Conference on Intelligent Transportation Systems*. IEEE, 2012, pp. 1055–1060.

[3] Z. Qin, X. Tang, Y. Jiao, F. Zhang, Z. Xu, H. Zhu, and J. Ye, "Ride-hailing order dispatching at didi via reinforcement learning," *INFORMS Journal on Applied Analytics*, vol. 50, no. 5, pp. 272–286, 2020.

[4] K. Park and I. Moon, "Multi-agent deep reinforcement learning approach for ev charging scheduling in a smart grid," *Applied Energy*, vol. 328, p. 120111, 2022.

[5] M. R. Alam and Z. Guo, "Co-optimization of charging scheduling and platooning for long-haul electric freight vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 147, p. 104009, 2023.

[6] P. Ruan, G. Wu, Z. Wei, and M. J. Barth, "A modularized electric vehicle model-in-the-loop simulation for transportation electrification modeling and analysis," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2021, pp. 1685–1690.

[7] M. J. Koshkouei, E. Kampert, A. D. Moore, and M. D. Higgins, "Impact of battery state of charge on in-situ power line communication within an intelligent electric vehicle," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 3855–3860.

[8] P. Wang and C.-Y. Chan, "Formulation of deep reinforcement learning architecture toward autonomous driving for on-ramp merge," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2017, pp. 1–6.

[9] J. Díaz-Ramírez, S. Zazueta-Nassif, R. Galarza-Tamez, D. Prato-Sánchez, and J. I. Huertas, "Characterization of urban distribution networks with light electric freight vehicles," *Transportation Research Part D: Transport and Environment*, vol. 119, p. 103719, 2023.

[10] A. J. Hoffman, P. Schutte, and S. J. Rabe, "Novel system for the monitoring of in-transit compliance of freight trucks," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 2220–2226.

[11] S. Pelletier, O. Jabali, and G. Laporte, "Charge scheduling for electric freight vehicles," *Transportation Research Part B: Methodological*, vol. 115, pp. 246–269, 2018.

[12] W. Lin, C. Zhu, W. Zhu, and S. Shen, "Charging scheduling strategies of cooperated car-hailing operating business for electric taxis," in *2021 International Conference on Wireless Communications and Smart Grid (ICWCSG)*. IEEE, 2021, pp. 407–413.

[13] C. Fiori and V. Marzano, "Modelling energy consumption of electric freight vehicles in urban pickup/delivery operations: analysis and estimation on a real-world dataset," *Transportation Research Part D: Transport and Environment*, vol. 65, pp. 658–673, 2018.

[14] Y. Liu, F. Wu, C. Lyu, S. Li, J. Ye, and X. Qu, "Deep dispatching: A deep reinforcement learning approach for vehicle dispatching on online ride-hailing platform," *Transportation Research Part E: Logistics and Transportation Review*, vol. 161, p. 102694, 2022.

[15] L. Cilio and O. Babacan, "Allocation optimisation of rapid charging stations in large urban areas to support fully electric taxi fleets," *Applied Energy*, vol. 295, p. 117072, 2021.

[16] Y. Zhao, Y. Xu, Y. Guo, and Q. Guo, "Reinforcement learning based optimal operation strategy for electric taxis," in *2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2)*. IEEE, 2020, pp. 2942–2947.