2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Human Preference-aware Rebalancing and Charging for Shared Electric Micromobility Vehicles

Heng Tan[1], Yukun Yuan[2], Hua Yan[1], Shuxin Zhong[3], Yu Yang[1]

***Abstract*— Shared electric micromobility has surged to a popular model of urban transportation due to its efficiency in short-distance trips and environmentally friendly characteristics compared to traditional automobiles. However, managing thousands of shared electric micromobility vehicles including rebalancing and charging to meet users' travel demands still has been a challenge. Existing methods generally ignore human preferences in vehicle selection and assume all nearby vehicles have an equal chance of being selected, which is unrealistic based on our findings. To address this problem, we design PERCEIVE, a human preference-aware rebalancing and charging framework for shared electric micromobility vehicles. Specifically, we model human preferences in vehicle selection based on vehicle usage history and current status (e.g., energy level) and incorporate the vehicle selection model into a robust adversarial reinforcement learning framework. We further utilize conformal prediction to quantify human preference uncertainty and fuse it with the reinforcement learning framework. We evaluate our framework using two months of real-world electric micromobility operation data in a city. Experimental results show that our method achieves a performance gain of at least 4.02% in the net revenue and offers more robust performance in worst-case scenarios compared to state-of-the-art baselines.**

***Index Terms*—Intelligent Transportation Systems, Human Factors and Human-in-the-Loop, Reinforcement Learning.**

I. Introduction

Shared micromobility has surged to a popular model of urban transportation. For example, shared electric scooters and bikes are growing steadily in the United States, increasing trips from 321,000 in 2010 to 112 million in 2021 [1]. As an alternative to traditional automobiles, shared electric micromobility allows users to travel in a more efficient and environmentally friendly way for short-distance trips, such as commuting from subway stations to working places [2], [3]. However, managing thousands of shared electric micromobility vehicles in a city remains a challenging problem, one of which is to rebalance and charge vehicles among different regions to meet spatial-temporally varying user demands [4], [5]. In this work, we use electric scooters (e-scooters) as an example to study the problem of jointly rebalancing and charging shared electric micromobility vehicles.

Various rebalancing and charging methods have been designed for shared electric micromobility vehicles [4]–[13]. They generally first set up a simulation environment based on historical vehicle usage data, such as pick-up/drop-off locations and energy consumption of trips. In this environment, a user randomly selects a nearby vehicle with adequate energy for the trip when multiple vehicles are available. Then, based on this environment, they employ either mixed integer programming methods [5], [6], [8] or sequential decision process-based methods such as reinforcement learning [7], [9], [10] for optimal rebalancing and charging strategies.

However, a key limitation of this setup is that it ignores human preferences in vehicle selection. We conduct a study on what features may affect human selection and show the results in Fig. 1. We found vehicle usage frequency emerges as a pivotal feature. This is, individuals might be more drawn to those with less historical usage (suggesting a potential fresher appearance). The vehicles' remaining energy is also an important feature. Ignoring human preferences in vehicle selection can lead to vastly different energy distributions among different regions. We show the energy distribution in five regions based on human preference-based selection and random selection in Fig. 2. We divide the whole city into equal-size grids (each is considered as a region) and summarize vehicles' remaining energy of each region in a boxplot. The figure shows that the vehicle energy distributions are much different under human preference-aware (HP) and random (w/o HP) vehicle selection. This could lead to entirely different strategies for rebalancing and charging, rendering existing approaches less effective in our context due to human preferences.

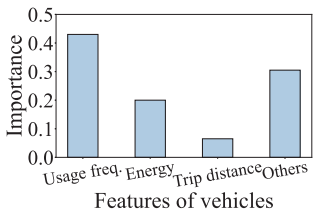

Fig. 1: Importance of vehicle features to user preference

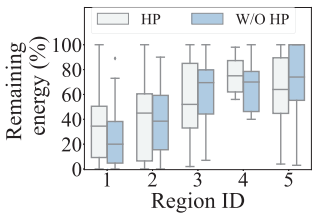

Fig. 2: Vehicle energy distribution under two vehicle selection strategies

To bridge the gap, we aim to design a human preference-

[1]Heng Tan, Hua Yan and Yu Yang are with the Department of Computer Science & Engineering, Lehigh University, Bethlehem, USA het221@lehigh.edu, huy222@lehigh.edu, yuyang@lehigh.edu

[2]Yukun Yuan is with the Department of Computer Science & Engineering, University of Tennessee at Chattanooga, Chattanooga, USA yukun-yuan@utc.edu

[3]Shuxin Zhong is with the Department of Computer Science & Engineering, Rutgers University, Piscataway, USA shuxin.zhong@rutgers.edu

This work was supported in part by NSF grants 2246080, 2318697, 2047822, and 1952096, and in part by the FY2024 and FY2025 Center of Excellence for Applied Computational Science competition in the University of Tennessee at Chattanooga. We also thank the anonymous reviewers for their valuable comments and feedback.

aware rebalancing and charging framework for shared electric micromobility vehicles. The opportunity for our work is that the shared electric micromobility system extensively records vehicle usage data, allowing us to model and predict users' preferences in vehicle selection. We have identified factors such as usage frequency and remaining energy that affect the choices of users, as shown in Fig. 1. In addition, the probability of a vehicle being selected in history also reflects how a vehicle will be selected in the future, suggesting some aspects that cannot be directly observed from vehicle usage data, such as how new a vehicle is. However, there are two challenges. First, the human preference modeling introduces uncertainty to vehicle scheduling (i.e., rebalancing and charging) due to model uncertainty. It is essential to quantify the uncertainty as it directly affects the effectiveness and efficiency of scheduling. Second, along with human preference uncertainty, scheduling models themselves, such as reinforcement learning, also present uncertainty [14]–[16] that may lead to unrobust scheduling performance in certain scenarios. It is challenging to deal with both scheduling uncertainty and human preference uncertainty simultaneously because of their intricate interplay.

We design PERCEIVE, a human preference-aware rebalancing and charging framework for shared electric micromobility vehicles with conformal vehicle selection prediction. First, we introduce conformal prediction [17] to quantify human preference modeling uncertainty, resulting in potentially selected vehicles with predefined precise confidence levels. Then, to incorporate human preference modeling into a scheduling framework considering both preference modeling uncertainty and scheduling uncertainty, we design a robust reinforcement learning framework to generate stable rebalancing and charging strategies by maximizing the expected reward under the worst cases. Specifically, we borrow the idea from adversarial reinforcement learning [18] and create two agents: a scheduling agent and an adversary agent. The scheduling agent maximizes the total revenues by generating better rebalancing and charging strategies, and the adversary agent minimizes the revenues by making the worst human vehicle selection. The scheduling agent and the adversary agent are trained alternatively by playing against each other until convergence.

In summary, the key contributions of this work are as follows:

- We are the first to consider human preferences in vehicle selection within the context of vehicle scheduling. We solve the problem of human preference-aware rebalancing and charging for shared electric micromobility vehicles.
- Technically, we design a robust adversary reinforcement learning framework for vehicle scheduling. The framework incorporates conformal prediction to quantify the uncertainty of human preference modeling and introduces two agents (i.e., a scheduling agent and an adversary agent) to learn a robust scheduling policy.
- By collaborating with a shared micromobility service provider, we evaluate our method based on real-world e-scooter usage data in a city. Our experiment results show that our method achieves a performance gain of at least 4.02% in net revenue and offers more robust performance in worst-case scenarios compared to state-of-the-art baselines.

II. Design

A. Problem Description

The problem of rebalancing and charging shared electric micromobility vehicles is that given the spatial-temporally varied user demand, the real-time locations of vehicles, and their remaining energy, we aim to decide how many vehicles in each region should be dispatched to other regions and how many of them should be charged, so as to meet users' demand while considering rebalancing and charging costs. Our goal is to maximize the total net revenue, which includes trip revenues but excludes rebalancing and charging costs.

Problem setting and notations: We partition the whole city into N equal-size grids as regions and regard the beginning of each trip as a time step based on the users' trip start time. For example, if there are T trips in a day, there will be a total T time steps. The vehicle's energy is divided into L levels, ranging from 0% to 100%. To describe the distribution of electric vehicles and their remaining energy spatially and temporally, we denote the number of vehicles of energy level l in the region i at time step t as $E_t^{i,l}$. To describe the spatial-temporally varied users' demand, we denote the number of users' trip requests from region i to region j with energy consumption level l from time step t to time step t' as $D_{t,t'}^{i,j,l}$. The trip revenue of satisfied users' requests from time step t to time step t' is defined as $R_{t,t'}^{trip}$.

Scheduling: When making rebalancing and charging strategies, the operator considers the current vehicles' energy distribution $E_t = \{E_t^{i,l} : \forall i \in N, \forall l \in L\}$ where N is the number of regions and L is the number of energy levels, and users' demand in the future h time steps $D_{t,t+h} = \{D_{t,t+h}^{i,j,l} : \forall i, j \in N, \forall l \in L\}$. We define reb_t and cha_t as the rebalancing and charging strategies at time step t, where $reb_t = \{reb_t^{i,j,l} : \forall i, j \in N, \forall l \in L\}$, and $cha_t = \{cha_t^{i,j,l} : \forall i, j \in N, \forall l \in L\}$.

Costs: After the operator gives the rebalancing and charging strategies $\{reb_t, cha_t\}$, the staffs drive trucks to reallocate the vehicles and charge them by swapping batteries, which causes rebalancing and charging costs, defined as C_t^r and C_t^c, respectively.

Objective: Our goal is to develop an optimal rebalancing and charging algorithm to maximize the total net revenue R:

$$\underset{reb_t, cha_t}{\operatorname{argmax}} \; R = R_{t,T}^{trip} - \sum_{t}^{T} (C_t^r(reb_t) + C_t^c(cha_t)). \quad (1)$$

B. Design Overview

We design a robust adversarial reinforcement learning framework for the rebalancing and charging shared electric micromobility vehicles, as shown in Fig. 3. This framework consists of three components: a scheduling component (the left part of Fig. 3 in green), an adversary component (the

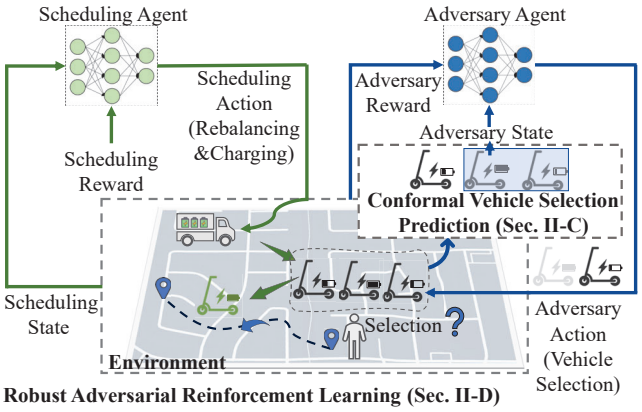

Fig. 3: An overview of RARL framework for rebalancing and charging electric micromobility vehicles with conformal vehicle selection prediction

right part of Fig. 3 in blue), and an environment component. (1) In the scheduling component, a scheduling agent determines how to rebalance and charge vehicles at specific times (e.g., midnight) to maximize net revenue. This agent takes the states of all vehicles as input and generates a scheduling policy as output. The policy is then implemented by trucks to carry out the rebalancing and charging operations. The input of this agent is the states of all the vehicles, and the output is the scheduling policy. The policy is then sent to trucks to perform rebalancing and changing. (2) In the adversary component, when a user initiates a trip request, we first input the state of all nearby vehicles into a conformal vehicle selection prediction module to predict a range of vehicles that are most likely to be selected. This range of the vehicles is then provided to the adversary agent, which perturbs the vehicle selection to minimize the net revenue (i.e., create worst-case vehicle selection). (3) Both components interact with the environment to obtain vehicle states and rewards (including scheduling rewards and adversary rewards). These two agents perform updates alternatively based on the corresponding rewards (the detailed training process is shown in Algorithm 1). In the following Section II-C and II-D, we first introduce how we model human preferences in vehicle selection and use conformal prediction to quantify the uncertainty of the model. Then, we introduce the robust adversarial reinforcement learning framework.

C. Conformal Vehicle Selection Prediction

To predict vehicle selection probabilities Y of each trip, we conduct extensive data engineering work and conclude the following important features, including the usage frequency, historical selection probability (i.e., the ratio of the usage frequency to the frequency of being potentially selected as nearby vehicles of a trip request), the geographical location, the remaining energy, trip distance, and the vehicle id. We test multiple methods, including SVM [19], XGBoost [20], and deep neural networks (DNN), and finally decide to use the DNN given its high accuracy (detailed comparisons in Section III). A typical DNN does not work in our problem, considering there may be different numbers of vehicles nearby for each trip (i.e., varied input and output sizes). To fix the size of the inputs and outputs of the prediction model, we consider all the vehicles as the input, while all the vehicles outside a specific range (e.g., 100 meters) and with remaining energy lower than the user's energy consumption request (we know the energy consumption based on the operational data) are padded as zero. We further add a mask layer before the activation function to avoid the impact of padded zeros on backward propagation. The output of the DNN is the selection probability of each vehicle nearby.

To quantify the uncertainty of vehicle selection prediction, we utilize conformal prediction [17] to predict the range of the vehicles that contain the actual selected vehicle (i.e., label) with a high confidence level, such as 95%, which is called prediction set. Formally, given a set of data (i.e., vehicles) $\mathbb{D} = \{(X^i, Y^i)\}_{i=1}^{M}$, we randomly divide it into a training set $\mathbb{D}_{\text{train}}$ and a calibration set $\mathbb{D}_{\text{cal}}$. All the samples are drawn exchangeable, corresponding to the assumption of the conformal prediction. We aim to construct a marginal distribution-free prediction interval $\mathbb{C}(X_{\text{cal}}) \in \mathbb{R}$ that is likely to contain the unknown response Y_{cal}. Therefore, given a confidence level α, we can obtain:

$$\mathbb{P}(Y_{\text{cal}} \in \mathbb{C}(X_{\text{cal}})) \geq 1 - \alpha. \tag{2}$$

Thus, the probability that the prediction set contains the correct label is almost exactly $1-\alpha$. We refer readers to [17] for more detailed definitions and processes of conformal prediction. By utilizing conformal prediction, we quantify the uncertainty of vehicle selection prediction. The conformal prediction results are then used in state transition and the adversary agent's action generation in the framework later.

D. Robust Adversarial Reinforcement Learning

Motivated by the advancement of vehicle scheduling [13] and robust adversarial reinforcement learning [18], we model the problem of rebalancing and charging shared electric micromobility vehicles as a Markov decision process expressed as a tuple $(\{\mathcal{S}^{sch}, \mathcal{S}^{adv}, \mathcal{A}^{sch}, \mathcal{A}^{adv}, \mathcal{R}^{sch}, \mathcal{R}^{adv}, \mathcal{P}, \gamma\})$. $\mathcal{S}^{sch}$ and $\mathcal{S}^{adv}$ are the continuous states of the scheduling agents and the adversary agent. $\mathcal{A}^{sch}$ and $\mathcal{A}^{adv}$ are the continuous sets of scheduling agents' and the adversary agent's actions. $\mathcal{P} : \mathcal{S}^{sch} \times \mathcal{S}^{adv} \times \mathcal{A}^{sch} \times \mathcal{A}^{adv} \times \mathcal{S}^{sch} \times \mathcal{S}^{adv} \rightarrow \mathcal{R}^{sch} \times \mathcal{R}^{adv}$ denotes the transition probability. $\mathcal{R}^{sch}$ and $\mathcal{R}^{adv}$ are the rewards of scheduling agents and the adversary agent. γ is the discounted factor. The definitions of these notations are as follows.

Agent: We assign a scheduling agent for each region, deciding the rebalancing and charging strategies for all the vehicles in the region, reducing the computational difficulty compared with a single scheduling agent for the whole system [21]. Then, we define an adversary agent for making the worst-case actions of human vehicle selection.

State: At time step t, the state of scheduling agent i contains the vehicle energy distributions in the whole city

and the predicted future users' energy consumption from time step t to $t+h$, denoted as $s_t^{sch,i} = \{E_t, D_{t,t+h}\}$. The state of the adversary agent at time step t is defined as $s_t^{adv,i} = \{E_t, D_{t,t+h}, V_t\}$. V_t denotes the vehicle selection probabilities of the trip at the time step t. The future users' energy consumption is predicted by a pre-trained prediction model used in [22]. Note that we do not attempt to design a new method to predict future energy consumption, considering it is not the focus of our work.

Action: Given the scheduling agent's state, the scheduling agent in region i at time step t decides the number of vehicles of energy level l needed to be rebalanced from region i to region j and the number of vehicles from them needed to be charged, denoted as $a_t^{sch,i} = \{reb_t^{i,j,l}, cha_t^{i,j,l}\}$. To make scheduling agents give periodic strategies, we define H as the time steps of the scheduling interval. In other words, all the scheduling agents do rebalancing and charging every H time steps. To disturb the vehicle selection of each trip, the adversary agent selects the vehicle that can be the worst-case selection based on its state at time step t, denoted as $a_t^{adv,i} = \{V_t'\}$.

Reward: All the scheduling agents collaboratively maximize the net revenue consisting of the trip revenue from satisfied trips, the cost of charging vehicles, and the cost for truck rebalancing:

$$r_t^{sch,i} = R_{t,t+H}^{trip} - \alpha \cdot \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{L} cha_t^{i,j,l} - \beta \cdot M, \quad (3)$$

where $R_{t,t+H}^{trip}$ is the total trip revenue from time step t to $t+H$. α and β are the weights. M is the total traveling mileage for truck rebalancing, which is provided by a truck-routing optimization method [22]. For the adversary agent, its reward is the revenue from unsatisfied users' demands, defined as $r_t^{adv} = R_t^{lost}$:

Transition probability function: It denotes the probability of state $s_t = \{s_t^{sch}, s_t^{adv}\}$ transferred to the next state s_{t+1} given the action $a_t = \{a_t^{sch}, a_t^{adv}\}$.

Discounted factor: Discounted factor γ represents the extent that agents pay attention to the future reward compared with the immediate reward, $\gamma \in [0, 1)$. If $\gamma = 0$, it indicates that the agent only cares about the immediate reward and learns the actions that cause the immediate reward.

Given the above setting, the objective of all the scheduling agents is to collaboratively maximize expected cumulative scheduling reward, which is denoted as $G_t = [\sum_{t=1}^{\infty} \gamma^{t-1} \sum_{i=1}^{N} R^{sch}(s_t^{sch,i}, a_t^{sch,i}) | s_1^{sch} = s]$. The Q-value of joint state s_t^{sch} and action a_t^{sch} under policy π_θ is denoted by: $Q^{\pi_\theta}(s_t^{sch}, a_t^{sch}) = E[\sum_{k=0}^{\infty} \gamma^k \sum_{i=1}^{N} R^{sch}(s_{t+k+1}^{sch,i}, a_{t+k+1}^{sch,i}) | \pi_\theta, s_t^{sch}, a_t^{sch}]$. The objective of the adversary agent is to maximize the expected cumulative adversary reward: $G_t = [\sum_{t=1}^{\infty} \gamma^{t-1} R^{adv}(s_t^{adv}, a_t^{adv}) | s_1^{adv} = s]$. The Q-value of state s_t^{adv} and action a_t^{adv} under policy π_δ is denoted by: $Q^{\pi_\delta}(s_t^{adv}, a_t^{adv}) = E[\sum_{k=0}^{\infty} \gamma^k R^{adv}(s_{t+k+1}^{adv}, a_{t+k+1}^{adv}) | \pi_\delta, s_t^{adv}, a_t^{adv}]$.

To optimize both the scheduling agents and the adversary agent, we use an alternating procedure to achieve Nash Equilibrium [18]. First, we collect the scheduling agents' trajectories $\{s_t^{sch}, a_t^{sch}, r_t^{sch}\}$ in the environment and improve their policies while keeping the adversary agent's policy constant. Then, we collect the adversary agent's trajectories $\{s_t^{adv}, a_t^{adv}, r_t^{adv}\}$ in the environment and improve its policy while keeping the scheduling agents' policies constant. We repeat this procedure until convergence. The Algorithm 1 outlines our method in detail.

Algorithm 1 RARL for rebalancing and charging of shared electric micromobility vehicles

Input: Environment; Vehicle selection probabilities of each trip; Stochastic policies of the adversary agent π_δ and scheduling agents $\pi_\theta = \{\pi_{\theta^i}\}$ $(\forall\ i \in N)$
Initialize: Learnable parameters in adversary and scheduling agents' policies; Distribution of initial state p
  for $n = 1$ to N_{iter} **do**
    Sample $s_1 \sim p$
    for $j = 1$ to N^s **do**
      Collecting N_{traj}^{sch} trajectories $\{s_t^{sch}, a_t^{sch}, r_t^{sch}\}$
      Updating the scheduling agents' policies π_θ
    end for
    for $j = 1$ to N^{adv} **do**
      Collecting N_{traj}^{adv} trajectories $\{s_t^{adv}, a_t^{adv}, r_t^{adv}\}$
      Updating the adversary agent's policy π_δ
    end for
  end for
  return π_θ, π_δ

III. Evaluation

A. Evaluation Methodology

Experiment setting: We conduct our experiments based on a two-month real-world shared e-scooter usage data [22], which contains vehicle IDs, vehicle locations, vehicle remaining energy, event types (e.g., trip start or trip end), and other relevant information. We divide the dataset into two parts: one month's usage data is used for training, and another month's data is used for testing. The vehicle's remaining energy is divided into 10 levels, ranging from 0 to 100%. Each region has a size of 800 meters $\times$ 800 meters, and the scheduling interval is 24 hours. The trip revenue is \$0.5 per minute based on the operator. The truck traveling cost per kilometer is set at \$2.422 based on the gas prices and truck fuel consumption (we ignore the labor fee to make it simple). The charging cost is valued at \$0.69 per e-scooter based on the electricity price and e-scooter battery volume.

Implementation: We implement our method and baselines with PyTorch 1.9.1, Python-mip 1.14.2, gym 0.21.0 in Python 3.7 environment and train it with 32 GB memory and GeForce RTX 3080 Ti GPU. A stochastic gradient descent optimizer is applied and the learning rate is 1e-4. The confidence level of conformal prediction is 0.9.

Baselines: We evaluate the performance of our model with the following five baselines:

TABLE I: Performance comparison of different approaches on the real-world data

Method	Trip Revenue ($)	Rebalancing Costs ($)	Net Revenue ($)	Daily Average Revenue ($)	Average Satisfaction Rate (%)
No Rebalance & Charging	592.59 (±5.96)	-	592.59 (±5.96)	83.3 (±0.85)	7.8 (±0.1)
State-of-The-Practice	5960.57 (±144.54)	1526.2 (±22.51)	4434.37 (±134.79)	851.51 (±20.65)	81.25 (±1.81)
Record [9]	6059.23 (±146.41)	913.88 (±21.55)	5145.35 (±133.42)	865.6 (±20.92)	81.6 (±1.94)
MADDPG [21]	5941.5 (±133.06)	1376.47 (±19.88)	4565.02 (±127.98)	848.79 (±19.01)	79 (±1.79)
RECOMMEND [22]	7015.63 (±114.87)	**486.60 (±16.91)**	6529.03 (±101.93)	1002.23 (±16.41)	94.5 (±1.54)
PERCEIVE	**7291.57 (±92.73)**	584.39 (±16.76)	**6707.19 (±85.46)**	**1041.65 (±13.25)**	**96.84 (±1.24)**

- **No Rebalance & Charging (NRC)**: There are no scheduling actions in the vehicle-sharing system.
- **State-of-The-Practice (SoTP)**: It represents the real-world scheduling policy based on a static charging threshold used by our platform collaboration.
- **MADDPG** [21]: It is a standard multi-agent RL framework to achieve cooperative or competitive relationships of multiple agents, which is commonly used in robotics and automation [23]–[25].
- **Record** [9]: It is a state-of-the-art electric carsharing rebalancing and charging algorithm based on the definition of the dynamic deadline for scheduling.
- **RECOMMEND** [22]: It is a state-of-the-art shared electric micromobility vehicle rebalancing and charging algorithm considering energy-informed demand.

Variants of our model: We conduct experiments considering the significance of different variants of our model:

- **Our model without conformal prediction (W/O CP)**: To verify the importance of conformal prediction on vehicle scheduling, we replace it with direct vehicle selection probability.
- **Our model without the adversary agent (W/O AA)**: To demonstrate the effectiveness of the adversary agent, we remove it and operate the shared electric micromobility system operation based on the conformal vehicle selection prediction result of each trip.

Metrics: For vehicle selection prediction, we utilize ACC, Recall, F-score, and Precision as metrics, which are widely used in prediction tasks [26]. To evaluate the performance of scheduling models, we use the monetary score (i.e., net revenue, trip revenue, and daily average revenue) and average satisfaction rate (i.e., the ratio of the number of satisfied trips to the number of total trips) as metrics. To evaluate the model robustness, we utilize the trip revenue and revenue decreasing rate (i.e., the decreasing rate of trip revenue under worst-case vehicle selections to that under human preference-aware vehicle selections) as metrics.

B. Overall Performance

Table I shows the overall performance of different methods. Our model achieves better trip revenue by at least 4.02% compared with state-of-the-art methods. The possible reason is that after the alternating optimization procedure with the adversary agent, the scheduling agents learn to generate robust rebalancing and charging strategies to handle the worst-case scenarios. Consequently, the scheduling agents' policy can satisfy more users' demands compared with other baselines. Table I also shows that PERCEIVE achieves more net revenue than other baselines even though its scheduling costs are higher than RECOMMEND. The possible reason is that the scheduling agents' policies always try to generate robust rebalancing and charging strategies to avoid low performance in the worst cases, while those worst-case situations seldom happen in the real-world shared micromobility system operation. As a result, it may cause unnecessary rebalancing and charging costs compared with RECOMMEND, whose goal is only to maximize the trip revenue while minimizing the scheduling costs. Besides the above metrics, the performance of PERCEIVE on satisfaction rate in Table I further supports our results.

C. The performance of conformal vehicle selection prediction

1) The performance of vehicle selection prediction: In our work, we use Deep Neural Network (DNN) to predict the vehicle selection probabilities based on the real-time vehicle information in the city. We conduct comparison experiments with other methods to evaluate the model performance, including XGBoost [20] and Support Vector Machine (SVM) [19]. Table II shows that the DNN-based method performs better than the other two methods.

TABLE II: Performance of different vehicle selection prediction models

Methods	ACC	Recall	F-score	Precision
XGBoost	0.742	0.692	0.551	0.685
SVM	0.653	0.628	0.499	0.619
DNN	0.855	0.731	0.583	0.726

2) The performance of conformal prediction: To evaluate the performance of conformal prediction, we use correctness coverage (the rate of covering the correct label) and the size of the prediction set (the preciseness of the conformal procedure) as metrics. The result is that the correctness coverage of the prediction set is 1.0, and the average size of the prediction set is 10.9. Therefore, conformal prediction can ensure that the prediction sets cover the 100% correct labels of vehicle selections in the test set, and the average number of potentially selected vehicles for each trip is 10.9.

D. Ablation Study

1) The effectiveness of conformal prediction: To demonstrate the effectiveness of conformal prediction, we conduct comparison experiments under different demand densities

compared to our model's variant W/O CP. Fig. 4 shows that even though the net revenue of two baselines is nearly the same when the user demand is low, the performance of PERCEIVE outperforms its variant W/O CP significantly by 10.94% when the user demand is high. The possible reason is that without conformal prediction, the model is trained toward a situation where some vehicles can still be selected even if they are very unlikely to be selected based on human preferences. In the real world, this situation rarely happens, which in turn decreases the model performance when we test on real-world data. This discrepancy is magnified when the demand is high.

2) The significance of the adversary agent: To prove the effectiveness of the adversary agent, we conduct comparison experiments with our model's variant W/O AA. Fig. 5 shows that optimized with the adversary agent, the scheduling agents learn to generate better rebalancing and charging strategies. As a result, it satisfies more users' demands with higher revenue compared to its variant W/O AA.

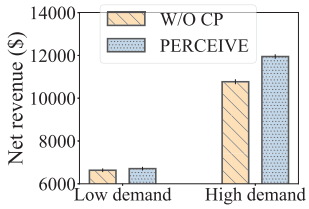

Fig. 4: The effect of conformal prediction

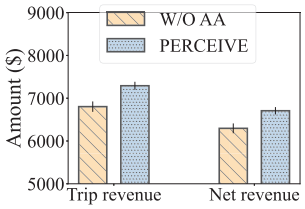

Fig. 5: The significance of the adversary agent

E. Robustness Analysis

To reflect the model robustness, we use four-week trips in the test set and conduct the comparison experiments with RECOMMEND under the worst-case vehicle selections that the adversary agent causes. Table III shows that the trip revenue of both methods decreases under the situation of worst-case vehicle selections caused by the adversary agent. However, PERCEIVE achieves a higher trip revenue and a lower revenue decline rate than RECOMMEND. This demonstrates that PERCEIVE can be more robust to different scenarios by introducing the adversary agent and human preference quantification.

TABLE III: The performance under the worst-case situation

Method	Trip Revenue ($)	Total Cost ($)	Revenue decline rate (%)
RECOMMEND	6623.02 (±109.49)	**565.09 (±17.03)**	5.6
PERCEIVE	**7037.15 (±90.61)**	660.46 (±18.54)	**3.1**

IV. Related Work

1) Vehicle Scheduling: Provided with real-time vehicle geographical locations and vehicle usage information, there is a substantial amount of work that focuses on addressing the imbalance problem between vehicle supply and user demand for various vehicle modes, such as (1) for-hire vehicles, including taxis [27]–[30], and e-taxis [5], [31]; (2) shared vehicles, including bikes [32]–[37], and e-scooters [13], [22], and e-cars [7], [8], [10]. The existing methods can be roughly categorized into two types from methodological perspectives: (1) Some researchers regard vehicle scheduling as a mixed integer programming problem with various constraints based on the spatial-temporal contexts and assumptions [5], [8], [33], [37]. (2) Other researchers model vehicle scheduling as a Markov Decision Processing and assign agents to give optimal rebalancing and charging (or rebalancing only) strategies through continuous interaction with vehicle operation environments [9], [30]–[32], [38]. However, the mentioned works do not consider human preferences for vehicle selection, leading to less realistic vehicle operation environments. Different from their works, we predict it as vehicle selection probability and incorporate it into an RL-based model.

2) Robust Reinforcement Learning: It aims to learn a policy that is robust to model errors in simulation and mismatch between training and testing scenarios [18]. The recent advances can be divided into two perspectives: (1) some researchers utilize constrained reinforcement learning and try to make agents maximize the worst-case expected reward while satisfying certain constraints to address the uncertainties in their models [14], [39], [40]. (2) Other researchers utilize adversarial reinforcement learning with a general framework of creating an adversary agent to play with a protagonist agent to maximize/minimize their rewards in the environment [18], [41], [42]. The adversary agent's reward is from the failure of the protagonist agent and its goal is to maximize such reward. As a result, the protagonist agent can be robust in different scenarios, trained with the adversary agent. Our work follows this framework and make adaptions considering the following difference. Compared to a free space of agent's actions in other works, the space of the adversary agent's actions (e.g., vehicle selection) is constrainted by human preferences. As a result, the action is conditioned on the human preference model. In our work, we use Conformal Prediction [43] to quantify the uncertainty of vehicle selection prediction and constrain the action space based on the uncertainty quantification. Then, we incorporate the constrained action space in state transition. Therefore, the adversary agent can make the worst-case action within the space of human-like vehicle selections.

V. Conclusion

In this work, we focus on the problem of human preference-aware rebalancing and charging for shared electric micromobility vehicles. We design a robust adversarial reinforcement learning (RARL) framework called PERCEIVE, which incorporates human preference in vehicle selections in the form of vehicle selection probabilities and contains an adversary agent to train the scheduling agent againt worst-case vehicle selection. To quantify the uncertainty of the vehicle selection prediction, we utilize conformal prediction and incorporate it into our RARL-based framework. The evaluation results show that PERCEIVE achieves an improvement of at least 4.02% in net revenue and more stable performance in the situation of worst-case vehicle selection compared with state-of-the-art methods.

REFERENCES

[1] C. Delgado, "Shared e-scooters can be sustainable—but there's a catch," https://www.popsci.com/environment/e-scooter-sustainability-micromobility/.

[2] S. Zhong, W. Lyu, D. Zhang, and Y. Yang, "Bikecap: Deep spatial-temporal capsule network for multi-step bike demand prediction," in *2022 IEEE 42st International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2022.

[3] W.-L. Shang, J. Chen, H. Bi, Y. Sui, Y. Chen, and H. Yu, "Impacts of covid-19 pandemic on user behaviors and environmental benefits of bike sharing: A big-data analysis," *Applied Energy*, vol. 285, p. 116429, 2021.

[4] S. He and K. G. Shin, "Dynamic flow distribution prediction for urban dockless e-scooter sharing reconfiguration," in *Proceedings of The Web Conference 2020*, 2020, pp. 133–143.

[5] Y. Yuan, D. Zhang, F. Miao, J. Chen, T. He, and S. Lin, "pˆ 2charging: proactive partial charging for electric taxi systems," in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2019, pp. 688–699.

[6] Y. Yuan, D. Zhang, F. Miao, J. A. Stankovic, T. He, G. Pappas, and S. Lin, "eroute: Mobility-driven integration of heterogeneous urban cyber-physical systems under disruptive events," *IEEE Transactions on Mobile Computing*, 2021.

[7] M. Luo, W. Zhang, T. Song, K. Li, H. Zhu, B. Du, and H. Wen, "Rebalancing expanding ev sharing systems with deep reinforcement learning," in *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, 2021, pp. 1338–1344.

[8] M. Zhao, X. Li, J. Yin, J. Cui, L. Yang, and S. An, "An integrated framework for electric vehicle rebalancing and staff relocation in one-way carsharing systems: Model formulation and lagrangian relaxation-based solution approach," *Transportation Research Part B: Methodological*, vol. 117, pp. 542–572, 2018.

[9] G. Wang, Z. Qin, S. Wang, H. Sun, Z. Dong, and D. Zhang, "Record: Joint real-time repositioning and charging for electric carsharing with dynamic deadlines," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 3660–3669.

[10] A. Bogyrbayeva, S. Jang, A. Shah, Y. J. Jang, and C. Kwon, "A reinforcement learning approach for rebalancing electric vehicle sharing systems," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[11] G. Guo and T. Xu, "Vehicle rebalancing with charging scheduling in one-way car-sharing systems," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[12] S. He and K. G. Shin, "Socially-equitable interactive graph information fusion-based prediction for urban dockless e-scooter sharing," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 3269–3279.

[13] H. Tan, Y. Yuan, S. Zhong, and Y. Yang, "Joint rebalancing and charging for shared electric micromobility vehicles with human-system interaction," in *Proceedings of the ACM/IEEE 14th International Conference on Cyber-Physical Systems (with CPS-IoT Week 2023)*, 2023, pp. 235–236.

[14] Z. Zhang, S. Han, J. Wang, and F. Miao, "Spatial-temporal-aware safe multi-agent reinforcement learning of connected autonomous vehicles in challenging scenarios," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 5574–5580.

[15] Y. Wang, F. Miao, and S. Zou, "Robust constrained reinforcement learning," *arXiv preprint arXiv:2209.06866*, 2022.

[16] S. He, S. Han, and F. Miao, "Robust electric vehicle balancing of autonomous mobility-on-demand system: A multi-agent reinforcement learning approach," *arXiv preprint arXiv:2307.16228*, 2023.

[17] A. N. Angelopoulos and S. Bates, "A gentle introduction to conformal prediction and distribution-free uncertainty quantification," *arXiv preprint arXiv:2107.07511*, 2021.

[18] L. Pinto, J. Davidson, R. Sukthankar, and A. Gupta, "Robust adversarial reinforcement learning," in *International Conference on Machine Learning*. PMLR, 2017, pp. 2817–2826.

[19] W. S. Noble, "What is a support vector machine?" *Nature biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.

[20] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[21] R. Lowe, Y. I. Wu, A. Tamar, J. Harb, O. Pieter Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," *Advances in neural information processing systems*, vol. 30, 2017.

[22] T. Heng, Y. Yukun, Z. Shuxin, and Y. Yu, "Joint rebalancing and charging for shared electric micromobility vehicles with energy-informed demand," *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2023.

[23] Y.-C. Liu, J. Tian, C.-Y. Ma, N. Glaser, C.-W. Kuo, and Z. Kira, "Who2com: Collaborative perception via learnable handshake communication," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 6876–6883.

[24] C. De Souza, R. Newbury, A. Cosgun, P. Castillo, B. Vidolov, and D. Kulić, "Decentralized multi-agent pursuit using deep reinforcement learning," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 4552–4559, 2021.

[25] R. Han, S. Chen, S. Wang, Z. Zhang, R. Gao, Q. Hao, and J. Pan, "Reinforcement learned distributed multi-robot navigation with reciprocal velocity obstacle shaped rewards," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 5896–5903, 2022.

[26] Z.-H. Zhou, *Machine learning*. Springer Nature, 2021.

[27] Y. Wang, H. Yin, H. Chen, T. Wo, J. Xu, and K. Zheng, "Origin-destination matrix prediction via graph convolution: a new perspective of passenger demand modeling," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 1227–1235.

[28] L. Bai, L. Yao, S. Kanhere, X. Wang, Q. Sheng *et al.*, "Stg2seq: Spatial-temporal graph to sequence model for multi-step passenger demand forecasting," *arXiv preprint arXiv:1905.10069*, 2019.

[29] L. Ling, X. Lai, and L. Feng, "Forecasting the gap between demand and supply of e-hailing vehicle in large scale of network based on two-stage model," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 3880–3885.

[30] S. Han, H. Wang, S. Su, Y. Shi, and F. Miao, "Stable and efficient shapley value-based reward reallocation for multi-agent reinforcement learning of autonomous vehicles," *arXiv preprint arXiv:2203.06333*, 2022.

[31] G. Wang, S. Zhong, S. Wang, F. Miao, Z. Dong, and D. Zhang, "Data-driven fairness-aware vehicle displacement for large-scale electric taxi fleets," in *2021 IEEE 37th International Conference on Data Engineering (ICDE)*. IEEE, 2021, pp. 1200–1211.

[32] L. Pan, Q. Cai, Z. Fang, P. Tang, and L. Huang, "A deep reinforcement learning framework for rebalancing dockless bike sharing systems," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 1393–1400.

[33] J. Li, Q. Wang, W. Zhang, D. Shi, and Z. Qin, "Dynamic rebalancing dockless bike-sharing system based on station community discovery." in *IJCAI*, 2021, pp. 4136–4143.

[34] J. Gu, Q. Zhou, J. Yang, Y. Liu, F. Zhuang, Y. Zhao, and H. Xiong, "Exploiting interpretable patterns for flow prediction in dockless bike sharing systems," *IEEE Transactions on Knowledge and Data Engineering*, 2020.

[35] M. Jiang, C. Li, K. Li, and H. Liu, "Destination prediction based on virtual poi docks in dockless bike-sharing system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 2457–2470, 2021.

[36] W. Wang, X. Zhao, Z. Gong, Z. Chen, N. Zhang, and W. Wei, "An attention-based deep learning framework for trip destination prediction of sharing bike," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4601–4610, 2020.

[37] R. Harikrishnakumar and S. Nannapaneni, "Smart rebalancing for bike sharing systems using quantum approximate optimization algorithm," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2021, pp. 2257–2263.

[38] B. Guo, S. Wang, Y. Ding, G. Wang, S. He, D. Zhang, and T. He, "Concurrent order dispatch for instant delivery with time-constrained actor-critic reinforcement learning," in *2021 IEEE Real-Time Systems Symposium (RTSS)*, 2021, pp. 176–187.

[39] S. Li and O. Bastani, "Robust model predictive shielding for safe reinforcement learning with stochastic dynamics," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 7166–7172.

[40] J. H. Gillula and C. J. Tomlin, "Guaranteed safe online learning via reachability: tracking a ground target using a quadrotor," in *2012 IEEE*

International Conference on Robotics and Automation. IEEE, 2012, pp. 2723–2730.

[41] X. Pan, D. Seita, Y. Gao, and J. Canny, "Risk averse robust adversarial reinforcement learning," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 8522–8528.

[42] C. Tessler, Y. Efroni, and S. Mannor, "Action robust reinforcement learning and applications in continuous control," in *International Conference on Machine Learning*. PMLR, 2019, pp. 6215–6224.

[43] G. Shafer and V. Vovk, "A tutorial on conformal prediction." *Journal of Machine Learning Research*, vol. 9, no. 3, 2008.