Taylor & Francis
Taylor & Francis Group

International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

Deep online recommendations for connected E-taxis by coupling trajectory mining and reinforcement learning

Wei Tu, Haoyu Ye, Ke Mai, Meng Zhou, Jincheng Jiang, Tianhong Zhao, Shengao Yi & Qingquan Li

To cite this article: Wei Tu, Haoyu Ye, Ke Mai, Meng Zhou, Jincheng Jiang, Tianhong Zhao, Shengao Yi & Qingquan Li (15 Nov 2023): Deep online recommendations for connected E-taxis by coupling trajectory mining and reinforcement learning, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2023.2279969

To link to this article: https://doi.org/10.1080/13658816.2023.2279969

Published online: 15 Nov 2023.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2023.2279969

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Deep online recommendations for connected E-taxis by coupling trajectory mining and reinforcement learning

Wei Tu[a,b,c], Haoyu Ye[d], Ke Mai[e], Meng Zhou[f], Jincheng Jiang[g], Tianhong Zhao[h], Shengao Yi[i,j] and Qingquan Li[a,b,c]

[a]Guangdong Key Laboratory for Urban Informatics, Shenzhen University, Shenzhen, China; [b]Shenzhen Key Laboratory of Spatial Information Smart Sensing and Services, Department of Urban Informatics, Research Institute for Smart Cities, School of Architecture and Urban Planning, Shenzhen University, Shenzhen, China; [c]MNR Key Laboratory for Geo-Environmental Monitoring of Great Bay Area, Shenzhen University, Shenzhen, China; [d]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China; [e]Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China; [f]School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, Guangdong, China; [g]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China; [h]College of Big Data and Internet, Shenzhen Technology University, China Shenzhen; [i]Department of Geography and Urban Studies, Temple University, Philadelphia, PA, USA; [j]Department of City and Regional Planning, University of Pennsylvania, Philadelphia, PA, USA

ABSTRACT

There is a growing interest in the optimization of vehicle fleets management in urban environments. However, limited attention has been paid to the integrated optimization of electric taxi fleets accounting for different operations as well as complex spatiotemporal demand dynamics. To this end, this study develops a real-time recommendation framework based on deep reinforcement learning (DRL) for electric taxis (E-taxis) to improve their system performance with explicit modeling of multiple vehicle actions and varying travel demand across space and over time. Spatiotemporal patterns of urban taxi travels are extracted from large-scale taxi trajectories. Spatiotemporal strategies are proposed to coordinate E-taxis' repositioning and recharging with optimized recommendation for next destinations and charging stations. A spatiotemporal double deep Q-network (ST-DDQN) is embedded in the DRL framework to maximize the daily profit. A prototype real-time recommendation system for E-taxis is implemented for the decision-making of E-taxi drivers and sensitivity analyses are carried out. The experimental results in Shenzhen, China suggest that the proposed framework could improve the overall performance. This study will benefit the promotion of connected E-taxis and the development of clean and smart transportation.

ARTICLE HISTORY
Received 23 April 2023
Accepted 1 November 2023

KEYWORDS
Geospatial big data; deep reinforcement learning; electric taxis; intelligent transportation systems; spatiotemporal patterns

CONTACT Meng Zhou zhoum89@mail.sysu.edu.cn

1. Introduction

Transportation is one of the main sources of fossil energy consumption and carbon emissions (Stocker 2014). In recent years, the electrification of transportation has become an international consensus to tackle global warming. The taxi industry, an important component of urban transportation, is receiving increasing attention for adopting environmental-friendly and increasingly efficient electric vehicles. With the advancement of vehicle technologies, Electric taxis (E-taxis) provide a promising alternative for efficient and energy-saving urban mobility services (Abduljabbar *et al.* 2019, Veres and Moussa 2020), relieving traffic congestion and reducing carbon emissions (Vazifeh *et al.* 2018), which improves the sustainability of urban systems. Many cities worldwide have planned or implemented taxi electrification schemes. Several cities and regions are now close to the ambitious goal of fully electrifying the taxi fleets, such as Beijing, London, and Shenzhen. For example, London realized the target of 9000 plug-in hybrid taxis on the road by 2020 (Hall *et al.* 2018). Notably, in 2019, Shenzhen became the first city in the world with a pure E-taxi fleet. It is widely anticipated that, through the electrification of transportation services, a reduction in energy consumption and pollution in urban areas could be achieved, leading to more sustainable and efficient urban transportation.

However, at present, the shortcomings of electric vehicles (EVs) are dissuading drivers and operators from adopting E-taxis. Compared with conventional internal combustion engine (ICE) vehicles, E-taxis have shorter driving ranges largely due to insufficient battery capacity (Kempton 2016, Tu *et al.* 2019, 2021). And limited by current charging technologies, E-taxis usually take hours to fully recharge, which is significantly longer than the time for refueling (Tu *et al.* 2021). These constraints reduce the on-road service time of E-taxis and consequently decrease the income of taxi drivers as well as the level of service of the system. These disadvantages are also amplified by inadequate charging facilities and low gasoline prices, thus hampering the incentive for adopting E-taxis. To disengage from such a predicament, it is necessary to enhance the viability and efficiency of E-taxis. The technological advancements of vehicular automation and connectedness exhibit a promising lead in tackling these issues. Through coordinated operations and intelligent vehicle control, the centralized management of potentially automated fleets could substantially enhance the system efficiency, safety and environmental sustainability. While fully automated vehicle fleets are still in development and not yet ready for the market, centrally managed service fleets have hit the road for some time now, providing on-demand services to millions of users worldwide (Xu *et al.* 2018). Under these schemes, how to better plan taxi operations would immensely affect their operational and economic efficiency, and thus influence the adoption of E-taxis and the performance of urban transportation.

The rise of information and communication technology (ICT) and intelligent transportation systems (ITS) has brought about a massive amount of data that reflect urban mobility and the functioning of the urban systems. Big data contain rich information on spatial dynamics and can yield new insights for addressing urban problems (Welch and Widita 2019) and facilitate research in urban informatics and spatial optimizations. For instance, large-scale taxi trajectories contain rich spatiotemporal information of urban travels, such as the distribution of taxi demand and travel distance, and the temporal rhythm of demands (Zheng *et al.* 2014). Based on these spatiotemporal knowledges, many studies have been conducted to improve the viability of E-taxis by optimizing different aspects of taxi operation, such as optimally locating charging stations based on spatiotemporal charging demand, balancing charging station utilization to reduce social cost, optimizing E-taxi recharging schedule, and recommending more profitable routes (Tu *et al.* 2021). On the other hand, various optimization methods have been proposed aiming at improving the efficiency of electric fleets under different service contexts such as carsharing (Folkestad *et al.* 2020) and on-demand shared automated EVs (Dean *et al.* 2022), typically with mathematical programming. In terms of the optimization of E-taxi fleets, most existing studies only focus on certain operations such as charging and vehicle routing. However, the daily cycle of an E-taxi is highly complex with many different operations ranging from picking up and delivering passengers and cruising to recharging and waiting in place for the next orders. At the same time, the actions and behaviors of E-taxis have cascading effects on their following operations (Tseng *et al.* 2019, Tu *et al.* 2021). For example, recharging during peak hours may result in a loss of orders and thus lower daily profit. Serving a passenger to a faraway destination can generate higher short-term income, but may lead to long cruising without orders. Therefore, it is essential to consider the long-term effect of different actions of E-taxis. Some existing studies focused on the decision-making process of E-taxis regarding cruising and recharging actions based on the expected income. But their approach cannot be well extended to a more practical scenario, where the waiting and repositioning of E-taxis also influences their profitability.

This study develops a real-time recommendation framework to improve the efficiency of the E-taxi operations in a hypothetical connected and centralized system and increase the profit of E-taxi drivers considering the dynamic patterns of travel demand for taxis and the various courses of action by E-taxis including selecting and serving orders, repositioning via cruising, recharging at charging stations, and maintaining static in place. A reinforcement learning (RL) approach is applied to optimize the actions of E-taxis, taking into account the spatiotemporal distributions of urban taxi travels extracted from massive taxi trajectories. Spatiotemporal strategies are formulated to coordinate E-taxis' repositioning and recharging by providing near-optimal locations or charging stations as the next destinations. A spatiotemporal double deep Q-network (ST-DDQN) embedded in the RL framework is proposed to maximize the daily profit. A real-time recommendation prototype system for E-taxis is developed for the decision-making of E-taxis. The experimental results in Shenzhen, China suggest that the system prototype could significantly improve the performance of the E-taxi system compared to selected baselines. The contributions of this study are summarized as follows:

- Spatiotemporal patterns of taxi travels are identified from massive taxi trajectories and leveraged to improve the online recommendation for E-taxis.
- A RL-based integrated recommendation framework incorporating the discovery of spatiotemporal patterns is developed for online destination recommendation for E-taxis.
- Extensive experiments showcase the capability of the proposed framework to improve the performance of the system with sensitivity analyses revealing the

impact of different influential factors including battery capacity, charging speed, and fleet size.

- It also adds to the research on spatial planning and optimization by combining spatiotemporally rich information from big trajectory data and optimization of taxi operations to enhance the performance and sustainability of urban systems.

The remainder of the paper is organized as follows. Section 2 reviewed related literature. Section 3 overviews the recommendation system and defines the problem. Section 4 describes the proposed real-time recommendation framework coupling spatiotemporal knowledge discovery and deep reinforcement learning. Section 5 reports the experimental results. Section 6 summarizes and discusses the main findings.

2. Literature review

2.1. Data mining and optimization with taxi trajectory

Vehicle trajectories acquired by global navigation satellite system (GNSS) or computer vision enable us to understanding city-wide human mobility patterns and improve urban travels with these patterns. Uncovering mobility patterns from massive vehicle trajectories has attracted much attention in relevant fields from urban computing (Zheng *et al.* 2014) to transportation (Zheng 2015). Based on origin-destination pairs or routes extracted from GNSS trajectories, a few studies have contributed to portraying urban mobility (Liu *et al.* 2015), revealing taxi travel patterns (Liu *et al.* 2010, Chen *et al.* 2021), and predicting travel demand (Castro et al. 2012). For example, Gong *et al.* (2016) inferred activity patterns based on Bayes' theorem using taxi traces in Shanghai to estimate trip purposes with results similar to the survey data on the spatiotemporal characteristics of urban taxi travels. Xiong *et al.* (2023) extracted traffic congestions from raw taxi trajectories and revealed the spatiotemporal propagation patterns of traffic congestion using Dynamic Time Warping and directed acyclic graphs.

Many efforts have been made to improve the viability of E-taxis in different aspects, such as locating E-taxi charging stations (Tu *et al.* 2016, Meng *et al.* 2020), scheduling E-taxi recharging events, balancing charging stations utilization, and route recommendations. Tu *et al.* (2019) used the GPS trajectories of ride-hailing drivers in Beijing to investigate the extent to which EVs could satisfy the demand for ride-hailing services. Their results suggest long-range vehicles and more extensive coverage of charging facilities are essential for the success of EV-based ride-hailing.

Optimizing recharging activities can directly improve E-taxis' efficiency with the more productive allocation of time for service. Considering the time-varying electricity cost, scheduling E-taxi fleets, or individual drivers' recharging can also lower the daily recharging fee and thus increase the profitability of E-taxis. E-taxi drivers' recharging choices tend to spatiotemporally aggregate around popular locations, e.g. shopping centers, airports, railway stations, which may induce excessive waiting. Coordinating E-taxis by providing them with suitable time slots and charging locations is found effective to balance charging station utilization and reduce the average waiting time. Much attention has also been paid to route recommendation for taxis, which influences the opportunities of finding passengers and the cost of cruising. Previous studies have developed route recommendation methods for ICE taxis with the objectives of maximizing the probabilities to find clients (Wang *et al.* 2015, Kumar *et al.* 2018) or increasing the chance of serving more profitable trips (Qu *et al.* 2014). These results, however, were for ICE taxis and not directly transferable to the context of E-taxis.

On the other hand, the daily cycle for an E-taxi is highly complex with different operations, e.g. picking up and delivering passengers, recharging, waiting, and repositioning, and one action might influence the performance in the following hours (Tu *et al.* 2021). For instance, recharging during peak hours may lead to longer waiting time for charging poles and thus shorter effective service hour later in the day. Such a cascading effect implies that smart E-taxi operation should not only consider the immediate benefit of an upcoming decision. Accounting for both routing and recharging choices, Tseng *et al.* (2019) proposed a recursive method to maximize the expected net income of E-taxi driver's sequential decisions. But their recharging strategy did not consider the potential waiting at the station. Tu *et al.* (2021) aimed at deciding on the best actions from repositioning and recharging based on the expected accumulative net revenue in the next hours. But both works left out the waiting and relocating behaviors of E-taxis, which were key to the cost-efficiency of the operation and the level of service in the form of waiting time reduction on the passenger side.

2.2. Reinforcement learning in transportation research

Reinforcement learning (RL) models learn from agents' interactions with the environment and gradually optimize agent decisions to maximize long-term rewards (Mnih *et al.* 2015). The main components of a reinforcement learning model include agents, environments, actions, states, and rewards. In essence, an agent takes an action and its current state changes in response. A reward from the environment is then generated as positive or negative feedback for the action, e.g. the revenue in the context of E-taxis. Then, the RL model updates the reward evaluation and the conditional probabilities of taking actions given agents' states, called the policy, based on the received reward. Through a series of trial-and-error processes, the RL model is able to identify the action with the maximum reward. Eventually, the agents can choose the best action in each decision step to obtain the maximum long-term benefit.

For its capability to handle complex interactions during the decision-making process, RL has been applied for the optimization of transportation operation and management in many areas including signal control, energy management, and delivery system. In particular, previous studies have adopted reinforcement learning and demonstrated its performance in optimizing taxi and ridesharing operations (Gao *et al.* 2018, Welch and Widita 2019, Wang *et al.* 2020, Singh *et al.* 2022, Liu *et al.* 2022, Xu *et al.* 2023). For example, Gao *et al.* (2018) utilized RL to estimate the optimal choices of ICE taxi drivers at different locations. Using the historical trajectories of taxis in Beijing, the experiment suggests that RL-based recommendation can effectively improve drivers' profits. Verma *et al.* (2017), on the other hand, proposed a model to optimize the next destinations and improve taxi drivers' revenue with real-world taxi trajectory data. Jindal *et al.* (2018) developed a RL-based system for carpooling to maximize efficiency and reduce traffic congestion. With a Spatio-Temporal Neural Network model to predict trip time from GPS data, their model obtained promising results in comparison to a fixed baseline policy. Similarly, Holler *et al.* (2019) focused on fleet management for ridesharing platforms using RL models with different reward specifications. Yu and Gao (2022), on the other hand, developed a batch offline RL approach to solve the taxi routing problem with the objective of profit maximization. Their results suggest that the offline model is more efficient than the online Q-learning method. However, the above-mentioned studies focused on ICE vehicles and the road routing decision. The complex interactions of taxis, passenger demands, and charging stations need to be accounted for in the RL framework to better optimize the operations of E-taxis.

In recent years, RL models have been applied to the optimization of operations and management involving EVs (Shi *et al.* 2020, Bogyrbayeva *et al.* 2022, Zhou *et al.* 2022). For instance, Shi *et al.* (2020) applied RL with by a decentralized learning and centralized execution strategy to operate a centralized EV fleet. Their experiment results indicate better performance over the benchmark method in cost reduction. Bogyrbayeva *et al.* (2022) proposed a nighttime rebalancing system for EVs with shuttle vehicles and a central controller. They developed a policy gradient approach with recurrent neural networks to minimize the time cost of relocation. More recently, Zhou *et al.* (2022) proposed a graph-based spatio-temporal multi-agent reinforcement learning (GMIX) framework with soft time windows to dynamically plan for EV routes to serve passenger requests. Their experiments suggested that the proposed model outperformed baseline methods on the selected metrics of service quality.

In summary, previous studies have focused on the optimization of taxi and ridesharing service through different approaches and with different targets. Recent studies have also increasingly utilized RL-based approach to improve the system performance and some have focused on the optimization of EV fleets. However, a comprehensive consideration of diverse actions and factors related to the operation of E-taxis is still lacking. In addition, the potential demands that the E-taxis may encounter after their actions would significantly affect the decision-making processes. Such spatiotemporally dynamic patterns of demands are often neglected in existing studies. The proposed approach in this study aims at the integrated modeling of the various types of E-taxi actions. we also account for the influence of the potential travel demands by mining from large-scale taxi trajectories to identify the spatiotemporal characteristics of taxi demands to facilitate the online recommendations for E-taxis.

3. Real-time recommendation system for E-taxis

3.1. System overview

The aim of this study is to provide an effective online recommendation system to improve the efficiency and the viability of E-taxi operations. Here, we hypothesize that the system of interest consists of an E-taxi fleet connected and managed by a central operator that resembles the current on-demand service providers and vehicles in the fleet operate fully cooperatively and undertake actions recommended by the central

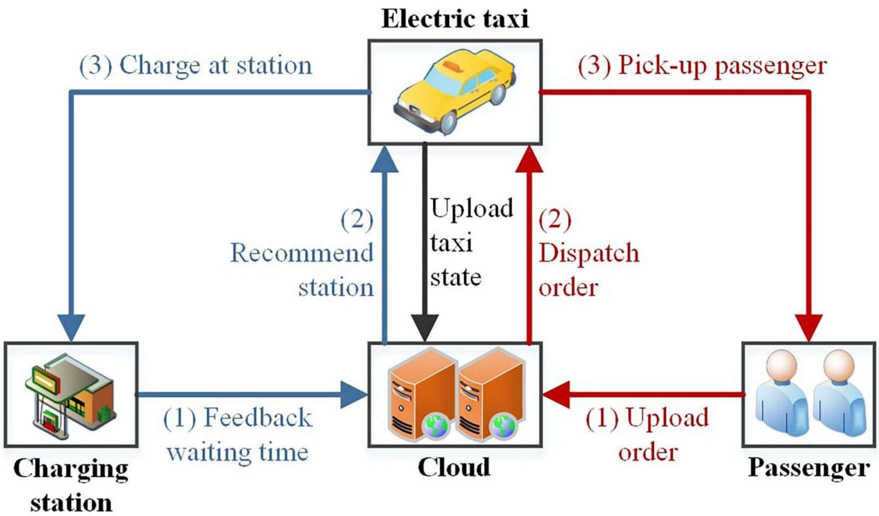

Figure 1. The real-time recommendation system for E-taxis.

operator. Hence, we present a real-time recommendation system that coordinates among taxis, passengers, and charging stations. An overview of the system workflow is shown in Figure 1. First, passengers submit their order requests via an online application and a cloud platform matches each request with the most suitable E-taxi. Considering the state of each idle E-taxi, the cloud platform will decide whether it should accept a request and pick up the passenger. If so, the taxi goes on to pick up and deliver the passenger to the destination. Otherwise, the taxi either remains on the road or goes to a charging station, depending on its remaining battery level. Simultaneously, charging stations keep updating and broadcasting real-time information, including the amount of charging piles and the number of vehicles in line. When the remaining battery level of an E-taxi cannot continue to sustain its operation, the cloud platform will recommend a charging station according to the queuing information. The idle E-taxis without assigned orders and sufficient battery, on the other hand, will reposition to nearby locations or wait in place for the next order based on the recommendation of the cloud platform. The maximum waiting time for passengers to be assigned a vehicle before cancellation is assumed to be 10 min.

The core of the recommendation system to coordinate and dispatch E-taxis is a deep reinforcement learning framework (ST-DDQN) incorporating the spatiotemporal patterns of taxi trips identified from massive trajectories. As mentioned, an important feature proposed in this study accounts for the potential travel demands at E-taxis' next locations to maximize the probabilities of receiving new orders after current actions so that the level of service and the system efficiency are improved. Two strategies are proposed based on these spatiotemporal characteristics in order to increase effective on-road service time and reduce potential passengers' waiting time. First, a cruising strategy is proposed to balance the spatial distribution of idle E-taxis and the travel demand. Second, a charging station recommendation strategy is formulated to reduce E-taxis' waiting time at charging stations and increase their chance of matching to passenger orders after recharging as soon as possible.

3.2. Preliminary problem statements

In this section, we first introduce relevant concepts and definitions and then briefly describe the problem of interest.

Definition 1. *A time step is a time instant at which all received orders to the operator are dispatched and the available E-taxis perform the recommended operations.*

At a time step, the cloud platform collects order requests and assigns them to E-taxis and the recommendation system decides if the order requests could be served. Requests that are not served will be dispatched at the next time step until they are finally served by E-taxis or cancelled by users. To capture the interactions among E-taxis, passenger demands, and charging services, at each time step, the input of the ST-DDQN model consists of an order request set $O = \{o_i\}$, an E-taxi set $E = \{e_j\}$, and a charging station set $C = \{c_k\}$. The output is the recommended decisions for E-taxis.

Definition 2. *A road junction l is the intersection point of roads in the network and represents the basic spatial unit of analysis.*

All locations of origins and destinations of the requests and locations of E-taxis and charging stations are geocoded to road junctions. The distance between two road junctions is calculated along the road network.

Definition 3. *An order request o_i is represented by a triplet $\{l_i^p, l_i^d, f_i\}$, where l_i^p and l_i^d represent the pick-up and drop-off locations of the trip, respectively, and f_i is the fare for an E-taxi serving the order.*

A customer submits an order request o_i to the system via a mobile device. All orders together represent the total demand. Upon receiving o_i, the cloud platform dispatches it to an E-taxi within 10 min. If o_i could not be assigned within that time, it is then regarded as an *unrealized order request*.

Definition 4. *An E-taxi e_j is represented by a tuple $\{l_j, b_j\}$, where l_j is the E-taxi's current location, and b_j indicates the remaining battery level of the vehicle.*

The state of an E-taxi depends on its previous state as well as its action at the last step. An occupied E-taxi will have its location and battery level updated in accordance to the destination and traveled distance or the location of the charging station. An idle E-taxi neither serving an order request nor charging at a station takes an action based on the recommended decision. Subsequently, its location will be updated and its remaining battery level will be deducted accordingly.

Definition 5. *A charging station c_k locates at a road junction and consists of several charging piles to provide recharging services.*

If an E-taxi e_j decides to recharge at a station c_k, it will submit a recharging request by sending its remaining battery level b_j to c_k. The E-taxi leaves c_k after its battery level is at or above 80% of full capacity. If all charging piles in c_k are occupied by E-taxis, newcomers must queue up and wait for available spots. Each charging station concurrently updates the minimum waiting time according to the received appointments to the control center.

Essentially, we focus on the design and modeling of a centrally operated E-taxi fleet in which orders are collected and dispatched via a central controller. E-taxis are coordinated by the central controller and they receive instructions about the next recommended actions including picking up and serving a passenger order, repositioning to another location, going to a charging station for recharging, and staying in place for potential orders. The central operator determines the action of each E-taxi based on the optimized policy learnt from the interactions of E-taxis and the dynamic demands for travel and charging service. The detailed learning approach is formulated in the next sections.

3.3. Markov decision process formulation

In this section, the decision-making process of our proposed system is formulated. We adopt an Markov decision process (MDP) to represent the interactions among E-taxis, travel demands and charging stations. An MDP models the decision process of intelligent agents for serialized decision optimization. The goal of an MDP is to find the optimal actions that maximize the expected long-term gains, often called rewards in the context of RL, of the agent taking the actions. In general, an MDP usually includes the following components: a state space $S = \{s\}$ that encompasses all states (conditions) of the agents, an action space $A = \{a\}$ that contains the actions for the agents to take, a state transition probability matrix $T = P(\bullet|s, a)$ to depict the transitions among states, and a reward function $R(s, a)$ characterizing the rewards that agents receive for their interactions with the environment. Upon the execution of an action, the state of an agent changes according to the transition probability given its current state and the action taken and the agent receives a reward from the environment. The objective then is to maximize the long-term cumulative reward over the whole process with a certain length or until a terminal state is reached. The MDP designed in this study mathematically formulates the hypothesized E-taxi service with each element described as follow.

State s: A state for an E-taxi is defined as $s = \{l_t, b_t, o_t\}$, where l_t and b_t denote location and remaining battery level at time step t, respectively, and o_t is the destination location if this E-taxi is serving an order.

Action a: The action set of an E-taxi contains four actions to choose from. The first is to serve an order by picking up and delivering the passenger (*Serve*). Another option is to go to a charging station and undergo recharging (*Charge*). The next action permits the E-taxi to cruise to a new location unoccupied (*Cruise*) and the final option is to stay in place and wait for a dispatched order (*Wait*).

As shown in Figure 2, at time t, E-taxi e_j at location l_t with remaining battery b_t chooses one of the available actions: (1) Picking-up passengers (*Serve*): if an order o_i is assigned to e_j, e_j can pick up the passenger and drive to the destination to fulfill the demand. (2) Charging (*Charge*): e_j drives to a charging station to recharge the battery. (3) Cruising (*Cruise*): e_j repositions without a passenger to the next location l_{t+1}. (4) Waiting at the current location (*Wait*): e_j stays in place waiting for the next order assignment. The energy consumed by the E-taxi for its action can be divided into the energy consumption for its displacement and the auxiliary loading energy

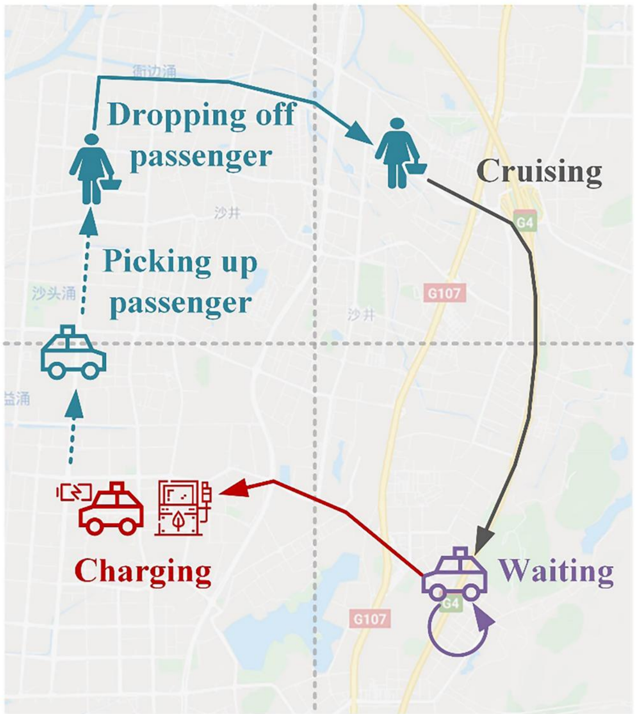

Figure 2. Actions of an E-taxi.

consumption. It is related to driving speed v, distance d and duration T and formulated as follows:

$$E = \beta(a_1 v^2 + a_2 v + a_3) \times d + \frac{\mu T}{60} \tag{1}$$

The aggressiveness parameter β captures the impact of driving behavior. Aggressive driving behaviors will significantly increase the rate of energy consumption and thus decrease the driving range, while mild driving behaviors can reduce energy consumption. Following Tseng *et al.* (2019), we define three types of driving behaviors: (1) aggressive behavior ($\beta = 1.2$); (2) normal behavior ($\beta = 1.0$); and (3) mild behavior ($\beta = 0.8$). The auxiliary load μ is highly related to the temperature and it is set to 1.5 kW in this study based on the historical records of temperature.

Reward r: The reward function determines the optimization goal of the model, which learns to obtain greater rewards through a trial-and-error process. According to the revenue and cost of the E-taxi operation, we set the reward function as the demand-adjusted net income of the taxi given an action. As shown in Table 1, each action has a specifically designed reward function to reflect the potential benefit and cost of taking the action, where r_{od} is the revenue for serving the order, E is the energy consumption of the action calculated using (1), P_{elec} is the unit cost of the consumed electricity or equivalent charging fee. ST_d is an adjusting parameter that reflects the spatiotemporal patterns at the potential destination after the chosen action identified from massive taxi trajectories, which will be elaborated on in Section 4. The adjusting parameter reflects the spatiotemporal characteristics of the demand for taxis

Table 1. Reward function.

Action a	Reward function r
Serve	$r_{od} \times ST_d$
Charge	$-E \times P_{elec} \times (1 - ST_d)$
Cruise	$-E \times P_{elec}/ST_d$
Wait	$-E \times P_{elec}/ST_d$

which are leveraged to enhance the effectiveness of the system recommendations. Here, we assume that *Cruise* and *Wait* action would not incur extra monetary cost (e.g. tolling or parking) beyond the consumption of energy in the simulated urban environments.

3.4. Recommendations for E-taxis

An E-taxi will make decisions sequentially over a period of time and obtain a series of rewards, which can be expressed as the reward sequence $[r_t, r_{t+1}, r_{t+2}, \cdots, r_T]$. We adopt a cumulative discounted reward function, $\mathbb{R}_t$, to estimate the long-term revenue of an E-taxi, as shown in Equation (2) below, which sums up the weighted future rewards of an E-taxi's sequential decisions over the horizon:

$$\mathbb{R}_t = \sum_{k=0}^{T-t} \lambda^k r_{t+k} \tag{2}$$

where $0 < \lambda \le 1$ is the discount factor that determines the degree to which the MDP looks into the future: a reward received after k time steps is only worth λ^k times of the immediate reward.

Formally, this E-taxi recommendation problem can be defined as follows: given an E-taxi in state s with action set A, find and recommend the optimal action to maximize its future reward $\mathbb{R}_t$. In this study, the DRL framework is incorporated with taxi trajectory mining to improve the long-term rewards of E-taxis and the overall efficiency of the E-taxi system.

4. Deep reinforcement learning based real-time recommendation

We develop a DRL model with ST-DDQN to optimize the sequential decisions of E-taxis with diverse actions. The ST-DDQN module receives real-time order requests from passengers and recommends the optimal action to E-taxi drivers. Here, we assume full cooperations and coordination by the E-taxi drivers. This framework is composed of an offline processing component, an online learning component, and an online recommendation component, as illustrated in Figure 3. In the offline processing component, historical taxi trip records are processed to extract spatiotemporal patterns of urban taxi travel demand. The cruising strategy and the recharging strategy are then developed partly based on the spatiotemporal characteristics of the demand. The ST-DDQN model, enhanced by a double deep Q-network, is developed and implemented. In addition, we build an experience filter to store different types of training samples, which makes the training process more reliable. Finally, the real-time recommendation system is set up to provide online recommendations after training.

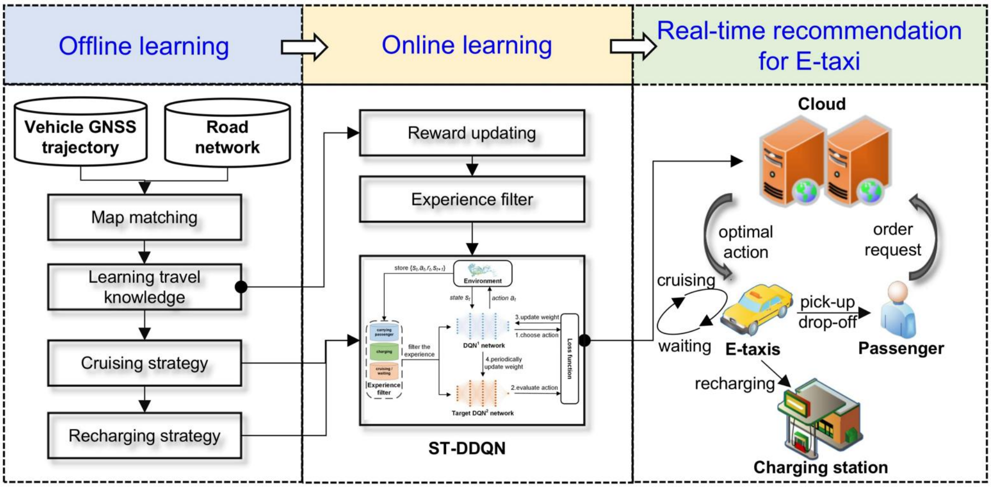

Figure 3. The workflow of the DRL-based real-time recommendation for E-taxi.

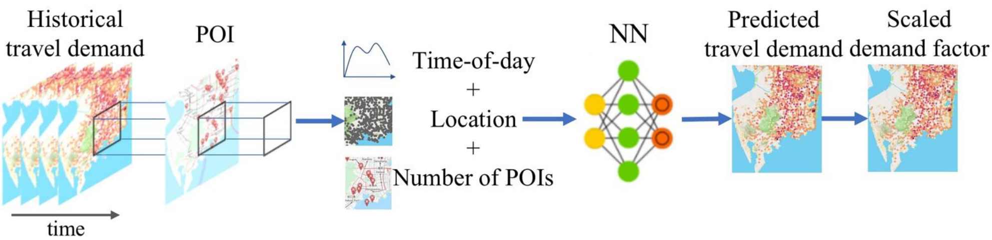

Figure 4. The neural network architecture for learning the distribution pattern of travel demand.

4.1. Strategies for taxi actions based on spatiotemporal demand dynamics

As mentioned above, the spatiotemporal patterns of the dynamic demands for taxi trips are leveraged to improve the decision-making process for the recommendation system. In this section, the data mining process to extract the spatiotemporal patterns of taxi travel demand and the subsequent strategies for relevant actions are elaborated on.

4.1.1. Extracting spatiotemporal travel patterns

The urban taxi travel demand is characterized with large spatial and temporal variations. It is essential to understand the spatiotemporal patterns of taxi demand before sending recommendations for E-taxis. Here, we utilize a neural network to learn the distribution of travel demand. Distribution patterns with strong spatiotemporal variabilities are obtained and referred to as the spatiotemporal demand factor *ST* hereafter, which is also mentioned in Section 3 and Table 1. Figure 4 shows the workflow of the proposed model in which a neural network (NN) module is built for travel demand estimation at the spatiotemporal unit level. The input of the NN module is a 3-D vector denoting the time, the location, and the activity potentials reflected by the number of nearby POIs. The output is then the predicted number of travel requests at different times and locations. Finally, the scaled spatiotemporal demand factor *ST* is obtained based on the dynamic demand prediction. The NN module consists of a three-layer multilayer perceptron (MLP) with different numbers of neurons at different layers. The hyperparameters are cross-validated, and those with the best performance are adopted. The numbers of layers and neurons in the three layers with the best performance are 64, 128, and 64 respectively. The NN model is trained with stochastic gradient descent and the mean absolute error (MAE) as loss function, as shown in Equation (3), where y_i and y_i^p are the true and predicted numbers of requests respectively and n is the number of spatiotemporal units. Based on the trained model, the *ST* factor at each time and location is obtained by normalizing the travel demand prediction.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_i^p\right| \tag{3}$$

Travel demand is highly space- and time-varying. The spatiotemporal demand factor *ST* describes the spatiotemporal variations in the distribution of travel demand for taxi. Based on historical taxi trip records, we extract the spatiotemporal patterns of travel demand and develop the NN module to estimate the dynamic demand and the ST factor to reflect the distribution of demand. The *ST* factor is normalized between 0.1 and 1 to indicate the relative amount of order requests made by customers at a specific location and time. It essentially describes the spatiotemporal variations in the distribution of travel demand. In our study, the value of *ST* indicates the potential number of upcoming orders at a specific location and time. The utility of the spatiotemporal demand factor is as mentioned in Section 3.3 and Table 1. If the E-taxi selects action *Serve* to pick up a passenger and serve the order, the reward for this decision is the revenue of the order r_{od} weighted by the spatiotemporal demand factor at the destination ST_d. When the E-taxi takes the action to serve an order that would end in an area with little travel demand, the reward for this choice will be low because of the low spatiotemporal demand factor and the small implied potential to pick up new orders at that location. The reward for action *Charge* is based on the cost of electricity C_{elec}. The spatiotemporal demand factor of choosing a charging station at location d, $1 - ST_{d,}$ associates a charging reward with the distribution of travel demand around that station. The reward for choosing a station with high travel demand is greater than that for choosing a station with low demand. The rewards obtained by the actions *Cruise* and *Wait* are also negatively correlated with the cost of electricity C_{elec}. Moreover, if the destination has a low travel demand, i.e. low ST_d, the E-taxi will obtain a low reward after the actions. Therefore, the strategies for destination choice for both *Charge* and *Cruise* are developed based on the estimated spatiotemporal demand factor.

4.1.2. Strategy for action charge

Compared to the refueling for the ICE vehicles, E-taxis take a longer time to recharge, which decreases taxis' service hour and thus total revenue. Efficient recharging recommendations will improve the efficiency of E-taxi operations and increase the revenue in the following time periods. Generally, a charging station with little travel demand around may have a short queue for recharging. But charging there may also induce a long cruising journey without order assignment afterwards. Therefore, we use the spatiotemporal demand factor to find the best charging station, c_r, as illustrated in Equation (4), rather than simply minimizing the sum of the waiting time to recharge, T_{d1}, and the travel time to the station, T_{d2}. Here, ST_d is the estimated normalized travel demand at the charging station as the destination of the recharging trip. This strategy not only reduces the total time cost for recharging but also considers the ease of being assigned new orders after.

$$c_r = \underset{d}{\operatorname{argmax}} \left(\frac{T_{d1} + T_{d2}}{ST_d} \right) \tag{4}$$

4.1.3. Strategy for action cruise

When the E-taxi decides to search for passengers on the road, the recommendation system will recommend a destination according to the current location, the remaining battery level, and the potential demand around the destination. Generally, the more demand at the destination of cruising, the higher the chance of being assigned new orders and thus higher revenue. Hence, the cruising strategy directs the E-taxis to locations with high levels of potential demand. Using the demand factor ST_d as the selection probability, the cruising strategy will recommend the location l_d with greater potential travel demand, as shown in Equation (5). Note that route planning is not a part of this study, and we assume that all E-taxis will choose the shortest paths to the recommended destination.

$$Pr(l_d) = ST_d \tag{5}$$

4.2. Spatiotemporal double deep Q-learning

DRL trains agents to learn the best decision by interacting with the environment. Here, we incorporate the spatiotemporal patterns of urban taxi travels into the RL framework to provide effective recommendations for E-taxis, as displayed in Figure 3.

4.2.1. Learning algorithm

Using the historical taxi trip records, the collected experience of the state-action pair is learned by the ST-DDQN model to update the value function $Q(s,a) = E[R|s;a,\pi]$, as shown in Equation (6). Policy π guides which actions should be taken given the states. The expected return $\mathbb{R}$ of one policy is determined by the value function. Finding the optimal action by $\max_{a_{t+1}} Q(s_{t+1}, a_{t+1})$, the action-value function $Q(s, a)$ can be estimated with tabular Q-learning:

$$Q(s,a) = Q(s_t,a_t) + \alpha(r_t + \lambda \max_{a_{t+1}} Q(s_{t+1},a_{t+1}) - Q(s_t,a_t)) \tag{6}$$

Here, α is the learning rate controlling the learning speed. The discount rate λ determines the present value of rewards received in the future. The tabular Q-learning algorithm uses a look-up table to store the Q function values, which is not suitable for cases with large or continuous states and action spaces. To better capture the complex characteristics of E-taxi operations, we use a deep neural network $Q(s,a;\theta)$ to

Table 2. Examples of Shenzhen taxi trajectory record.

Num	Pick-up time	Drop-off time	Pick-up location	Drop-off location	Trip distance (km)	Fare (CNY)
1	7:00	7:06	5724	4262	2.69	10.55
2	8:10	8:13	803	1614	1.45	6.8
3	17:15	17:20	2149	3726	2.25	9.73

approximate the Q function, which is called the deep Q-learning network (DQN), where θ represents the parameters of the neural network (Mnih *et al.* 2015).

Since the spatiotemporal characteristics of the taxi demand are incorporated into the reward function (as in Table 1), the value function can better capture the dynamics of demand at detailed temporal and spatial scales, which are potentially of great value to the optimization of E-taxis' decision-making processes. As such, the learning processes of agents in the RL model are supported by rich information from the massive historical taxi trip data (Table 2).

To improve the stability of the training process, we use double-DQN (DDQN) (Hasselt *et al.* 2016), where DQN^1 with parameters θ^1 is used to select the optimal action, and the target neural network DQN^2 with parameters θ^2 evaluates the value of the chosen action. By updating the parameters of DQN^2 periodically, we improve action value evaluation and the action selection. The neural network convergence target Y is kept stable for a certain period. Therefore, the target Y is modified as follows:

$$Y^{DDQN} = rt + \lambda Q(s_{t+1}, arg\max_{a_{t+1}} Q(s_{t+1}, a_{t+1}; \theta^1); \theta^2) \tag{7}$$

Using the minibatch update strategy, the DRL-based E-taxi operation optimization problem is solved through back propagation with the following loss function:

$$L_i(\theta_i) = \mathbb{E}_{(s,a)\sim\rho(\cdot)}\left[\left(Y_i^{\mathrm{DDQN}} - Q(s,a;\theta_i)\right)^2\right] \tag{8}$$

4.2.2. Experience filter

The ST-DDQN model updates its strategy by interacting with the environment, as shown in Figure 5. Here, the E-taxi driving environment is built by simulating city-scale E-taxis, passengers, and services at charging stations. As stochastic simulations lack real world experience, we use historical taxi trip data to simulate real-world cases for training, wherein the E-taxis pick up and deliver passengers, cruise, and recharge at the charging stations and obtain the corresponding rewards. The state-action-reward pairs are stored in memory for the training process. The E-taxi agent periodically takes a minibatch of samples from the experience pool to update the deep neural network. This strategy is called experience replay and has been widely used in reinforcement learning models.

However, urban traffic varies significantly across space and time. Previous studies usually store all the state-action-reward training samples in an experience pool (Wang *et al.* 2020, Verma *et al.* 2017). Over the episodes, the pool will be occupied by the most frequent state-action-reward pair. The E-taxi agent will repeatedly learn how to estimate the returns of these frequent state-action pairs while ignoring other infrequent pairs. This can cause the model to make incorrect decisions under certain circumstances. Here, we solve this imbalance problem by using the experience filter for

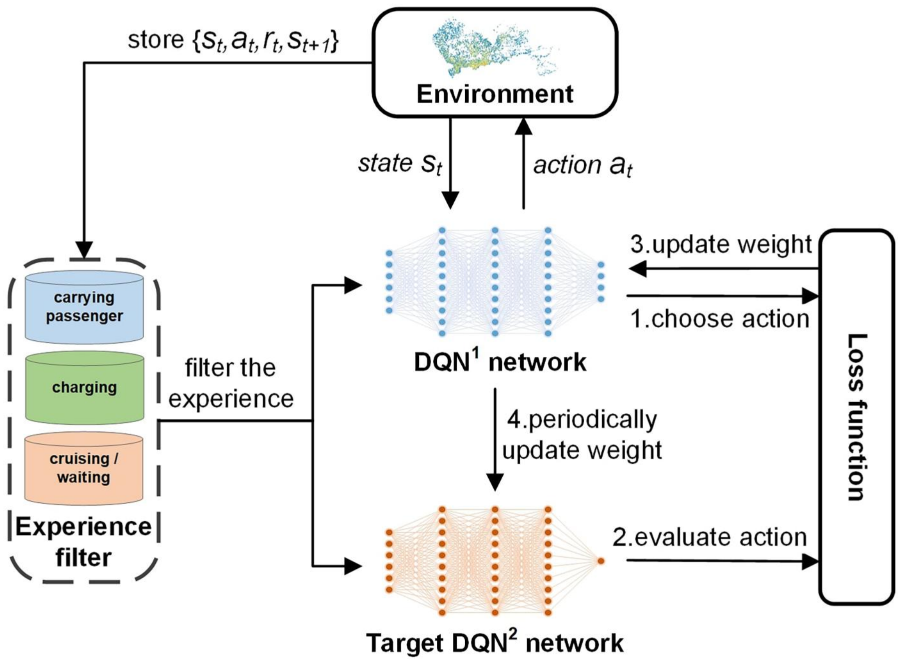

Figure 5. The ST-DDQN structure.

different state-action pairs. As shown in Figure 6, the training samples for different actions are stored in separate experience pools. Different numbers of samples are selected based on a roulette strategy and pre-defined proportions of different samples are obtained. As such, it can prevent unbalanced sample production from interfering with the training process, thus improving the capability of the network to predict the reward accurately and representatively.

5. Experiment and results

The recommendation system prototype is developed to optimize the E-taxi operations in Shenzhen, China, and evaluated for its performance. We first set-up the experiments following the real-world E-taxi operations. We then examine the performance of ST-DDQN by comparing it with selected baseline methods. Finally, we evaluate the sensitivities of ST-DDQN by changing several parameters to simulate different E-taxi operation scenarios.

5.1. Experiment setting

We used taxi trajectory data for Shenzhen, China to conduct the experiment, which were collected between March 1st and June 30th, 2016. The dataset includes the time, location, and occupied status of the 21,642 fuel-based taxis. We removed anomalous taxi trips, e.g. trips with zero travel time or travel distance, missing pick-up/drop-off locations, or destinations outside of Shenzhen. The dataset is split into a training set of 16,565,112 records generated between March 1st and May 31st, 2016 and a test set

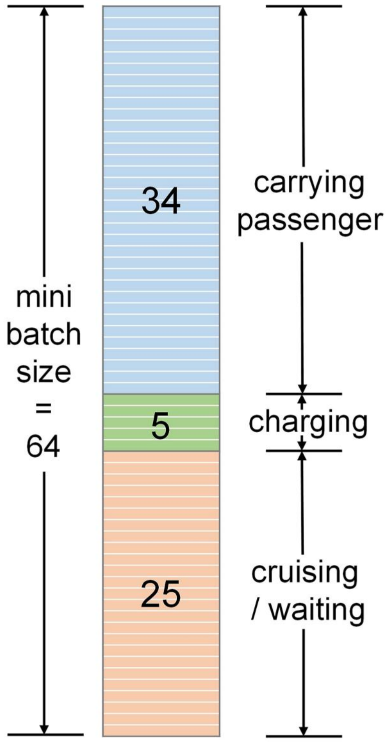

Figure 6. Experience filter structure.

of 1,167,078 records in the first week of June 2016. The road network of Shenzhen was crawled from OpenStreetMap and all pick-up and drop-off locations are adjusted to the nearest road junctions. The information of 97 charging stations was collected from the Shenzhen Transport Bureau. The charging pile counts and the rate of charging piles were also collected. 6000 BYD E6 E-taxis, a currently popular model, were designated as our agents. The battery capacity of BYD E6 is 60 kWh and its driving range is around 294 km. The charging rate is set to 40 kW. According to the current policy, the electricity price is 0.8 CNY/kWh.

The performance of the ST-DDQN is examined in four aspects: the ratio of served demand and evaluation metrics for charging, cruising, and waiting. To evaluate the performance, we compared the results from our model with those of the following four baseline scenarios, among which the first three are heuristics selected as the conventional baseline methods and the vanilla DDQN without spatiotemporal knowledge is selected as the state-of-the-art DRL baseline method. Initially, standard DQN was also implemented but it failed to converge, possibly due to the overestimation of state-action values and the lack of experience filter, making it vulnerable to the disturbance of unbalanced samples.

1. **DistFirst**: The E-taxi goes to the nearest charging station to recharge when the remaining battery power is below 20%.
2. **TimeFirst**: The E-taxi goes to the charging station with the shortest waiting time to recharge when its remaining battery power is below 20%.

Table 3. Performance of ST-NN and the baseline method for spatio-temporal knowledge mining.

Method	MAE	MRE
ST-NN	0.52	0.29
Historic Mean	1.98	0.54

3. **ICE**: Test results are also compared against those from scenarios with Internal combustion engine (ICE) taxis under identical conditions. ICE-based taxis refuel faster and have more time to serve passenger demands.
4. **DDQN**: Unlike the previous baseline methods, which are predefined heuristics, the DDQN method leverages reinforcement learning to adapt and optimize decisions based on the underlying environment. It employs a dual-network structure, comprising a primary network for action selection and a target network for value assessment, to overcome issues of action overestimation commonly encountered in Q-learning-based algorithms. The standard DDQN model without the strategies based on spatiotemporal knowledge from trajectory data is implemented and a baseline in this study.

5.2. Spatiotemporal patterns of taxi trips

Using the historic taxi trajectories in Shenzhen, spatiotemporal patterns of taxi travel demand are identified. We use the datasets collected from March to May and the first week of June as the training and the testing subsets respectively. Cross-validation with different hyperparameters was conducted to obtain the best performance of the NN model. The time window of the prediction is set to 15 minutes, in accordance with the average duration of taxi trips.

We compared the results of the NN model with the historic mean value prediction. The mean absolute error (MAE) and the mean relative error (MRE) between the estimated number of taxi orders and the real number of orders were calculated. Table 3 reports the results of the NN model and the historic mean method. It demonstrates that the NN model predicts future taxi order with an MAE of 0.52 and an MRE of 0.29, indicating large improvements over the simple historic mean method with an MAE of 1.98 and an MRE of 0.54. The NN model achieves an approximately 74% improvement in MAE and a 47% improvement in MRE. Therefore, this module provides a good prediction of city-wide taxi demand.

Figure 7 illustrates the temporally-varying estimated taxi demand. It shows that the NN method can capture the dynamics of the city-scale taxi demand reasonably well. From the spatial perspective, we normalize the predicted taxi orders at certain times of day and locations, as shown in Figure 8. It demonstrates that taxi orders tend to concentrate around the city center and transportation hubs such as the high-speed railway station and the airport.

5.3. The performance of ST-DDQN

By coupling the spatiotemporal information on taxi travel demand, the ST-DDQN module was trained to obtain the optimal operation policy for each E-taxi. Figure 9 shows

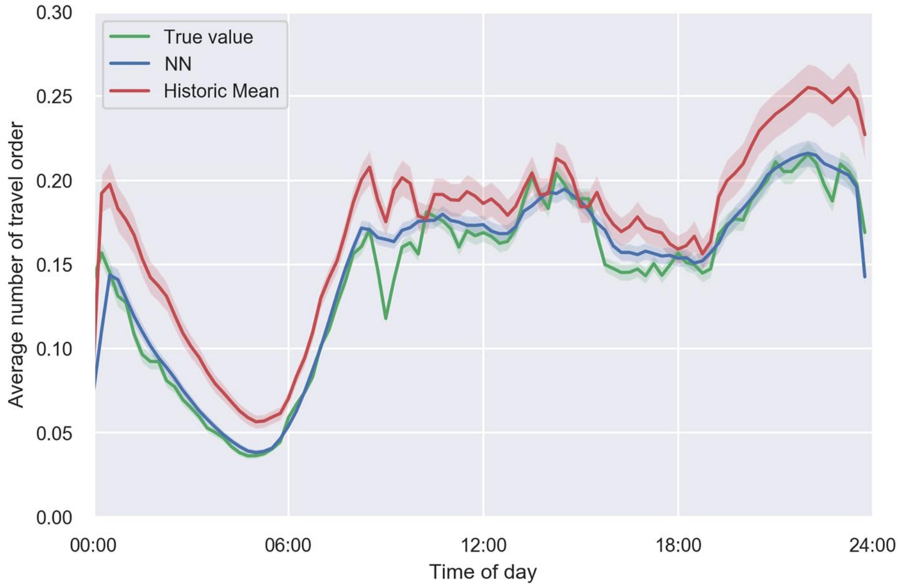

Figure 7. The prediction of travel demand by NN model and Historical mean method.

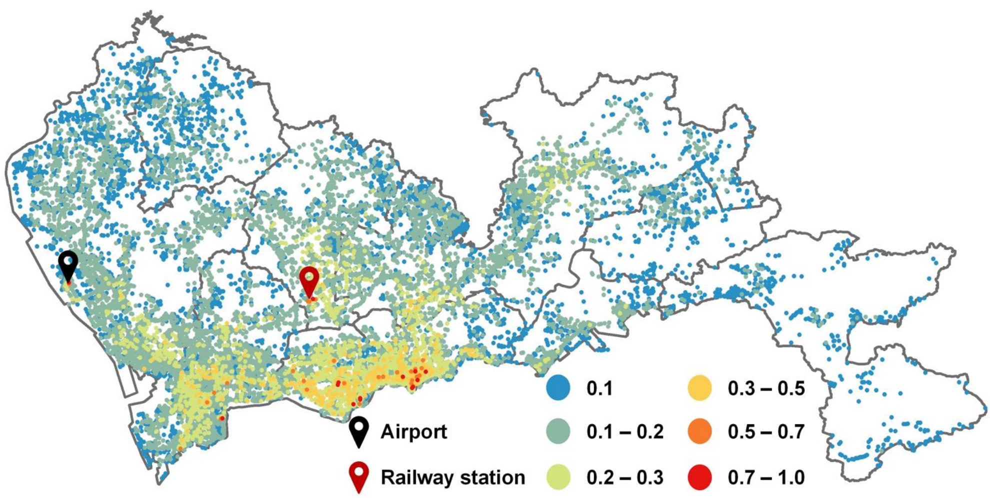

Figure 8. Spatial distribution of the *ST* factor at 09:00.

the training processes of ST-DDQN both with and without an experience filter. The results demonstrate that the filtering strategy with pre-defined proportions of samples significantly improves the ST-DDQN performance. With the experience filter, the total reward is gradually improved and finally converges at around 110 thousand after 91 episodes. While the model without experience filter converges at around 45 thousand.

In addition, we evaluated the performance of the trained ST-DDQN with the trip data in the first week of June 2016. Table 4 describes the detailed ST-DDQN results, including the percentage of served orders, various indicators for passenger delivery, charging, cruising, and waiting. These results are compared against those obtained

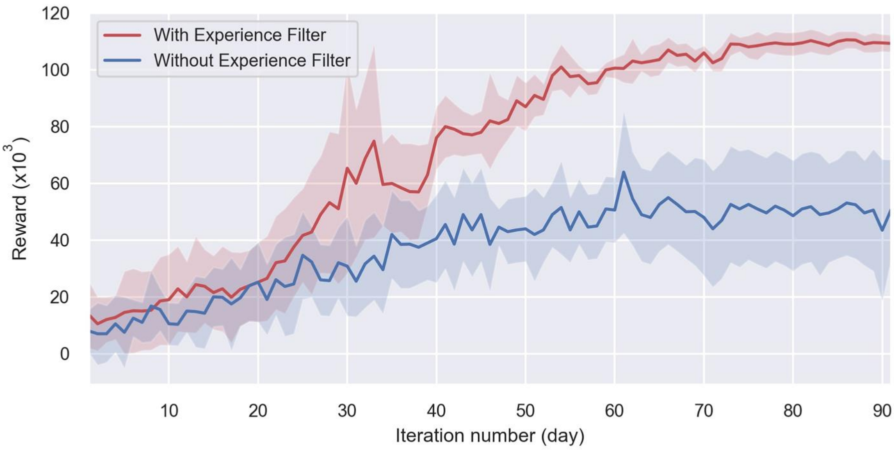

Figure 9. The improvement of ST-DDQN model with experience filtering.

Table 4. The performance of the ST-DDQN and four baseline methods.

		Passenger delivery			Charging		Cruising		Waiting
	Percent of taxi trips served	Net revenue (CNY)	Distance (km)	Time (hour)	Time (hour)	Waiting time (hour)	Distance (km)	Time (hour)	Time (hour)
ST-DDQN	97.57%	894.98	267.10	8.79	3.54	0.67	366.65	10.50	0.50
DistFirst	44.30%	385.52	113.63	3.74	1.00	17.13	69.49	2.35	–
TimeFirst	91.36%	846.58	255.19	8.42	3.39	0.39	340.11	11.80	–
ICE	95.91%	880.07	262.59	8.78	–	–	450.85	15.22	–
DDQN	52.83%	461.89	143.74	4.81	3.40	0.43	436.26	14.56	0.80

from the baseline methods. The findings reveal remarkable improvements achieved with the implementation of the ST-DDQN module. It shows that, with the ST-DDQN module, an E-taxi can earn 894.98 CNY (around 130 US dollar) per day on average with 8.79 h of service, covering 267.10 km of distance traveled. Because of the inherent spatiotemporal dynamic characteristics of taxi demand, E-taxis averagely spend 10.50 h to travel for 366.65 km to search for passengers. Meanwhile, E-taxis spend 3.54 h to recharge with an average waiting time of 0.67 h. These results indicate significant improvements over the baseline scenarios with larger revenue, longer service distances, and short waiting time for charging on par with the time minimization strategy. While the two minimization strategies, by design, outperform in either the distance or the time metric, overall, they fall behind the ST-DDQN model. The DistFirst method would induce significant congestions at popular charging stations, resulting in long waiting time for recharging service and a low level of trip completion rate. The TimeFirst method also results in fewer trips served and therefore less revenue. The ICE scenario is also compared against the proposed model and, despite its great advantage in refueling speed, it is still outmatched by the ST-DDQN model both in terms of revenue and cruising efficiency. The results of the baseline DDQN models report a 52.83% trip coverage and a net revenue of 461.89 CNY, albeit with longer cruising, charging, and waiting times. The results collectively emphasize the superior performance of the ST-DDQN approach, particularly with the spatiotemporal patterns of potential demand accounted for, both in terms of revenue generation and operational

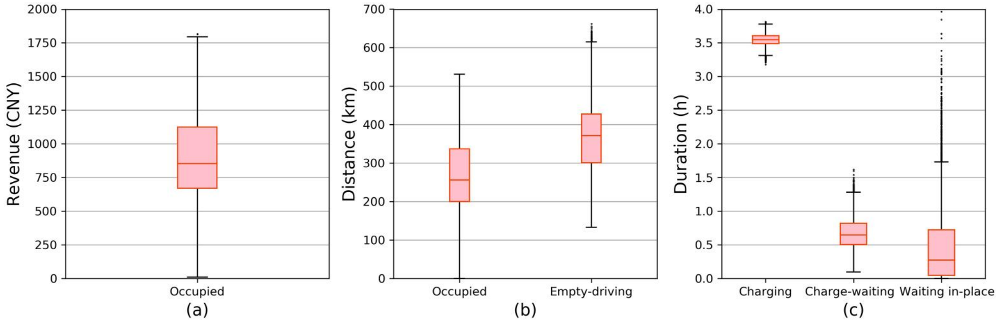

Figure 10. Distribution of the ST-DDQN E-taxis': (a)revenue; (b) occupied and cruising distances; (c) duration of charging, waiting for charging and waiting in place.

efficiency, underscoring its suitability for optimizing connected E-taxi charging strategies.

Figure 10 further illustrates the details of the results. It demonstrates that more than 90 percent of E-taxis have a daily net revenue of over 520 CNY (around 75 US dollar), while some E-taxis can even earn more than 1500 CNY (around 215 US dollar) per day. The daily occupied distance distributes similarly. All E-taxis cruise more than 120 km per day to search for order requests. Because of the limited driving range, E-taxis have to recharge two or three times per day. Figure 10(c) displays the distribution of the recharging time and the waiting time at charging stations. It shows the E-taxis wait at the stations for between 0.1 and 1.6 h due to the scarcity of charging stations. To reduce empty cruising on the road, E-taxis sometimes wait in place for the next orders. The waiting time varies from 0 to 2.5 h as Figure 10(c) displays.

5.4. Spatial distribution of E-taxi services

Figure 11 illustrates the spatial distribution of the ratio of unserved orders and the waiting time at charging stations. Because of the effective assignment by the ST-DDQN model, only about 2.43% of the orders are unmet. The percentages of unmet demand in most areas are well below 1%. Regardless, relatively higher amounts of unmet demands appear at the east and the north of the city. With the help of spatio-temporal patterns of demand, E-taxis effectively fulfill the taxi demand in urban areas. Figure 11(b) displays the average waiting time for recharging at charging stations. It suggests E-taxis wait, on average, for less than 15 min at all but 5 stations. These results imply that the ST-DDQN framework tends to recommend stations with higher travel demand to reduce the waiting time before and the cruising distances of E-taxis after recharging.

5.5. Sensitivity analysis

To further investigate the proposed framework in regard with its response to the varying hyperparameters, sensitivity analyses are conducted to assess the influence of several selected factors on the performance of the system.

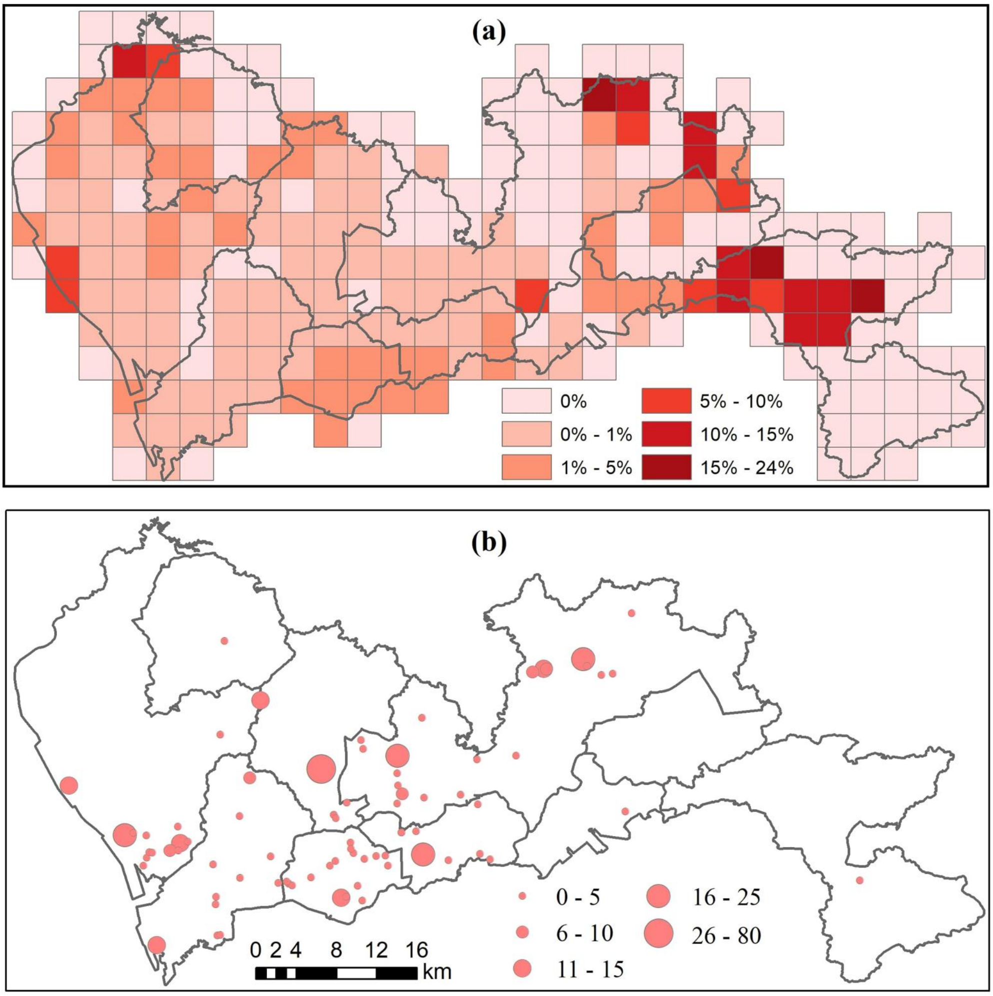

Figure 11. Spatial distribution of (a) the ratio of unmet demands; (b) waiting time at charging station (in minutes).

5.5.1. Impact of battery capacity

We first evaluated the effect of battery capacity by varying it between 40 kWh and 80 kWh. Figure 12(a) shows the changes in average daily net revenue and the percentage of unmet travel demand under different battery specifications. With increasing battery capacity, E-taxis supported by the ST-DDQN model improve their average daily revenue from 877 CNY (around 127 US dollar) to 897 CNY (around 130 US dollar). In terms of the unrealized demand, the ratio decreases from to 4.64% to 2.39%. Combined with results in Table 4, it indicates that the increase of battery capacity from 60 kWh to 80 kWh only improves the E-taxi service marginally and electric vehicles with the capacity in the contemporary time, given effective dispatching and management, are adequate for the taxi service.

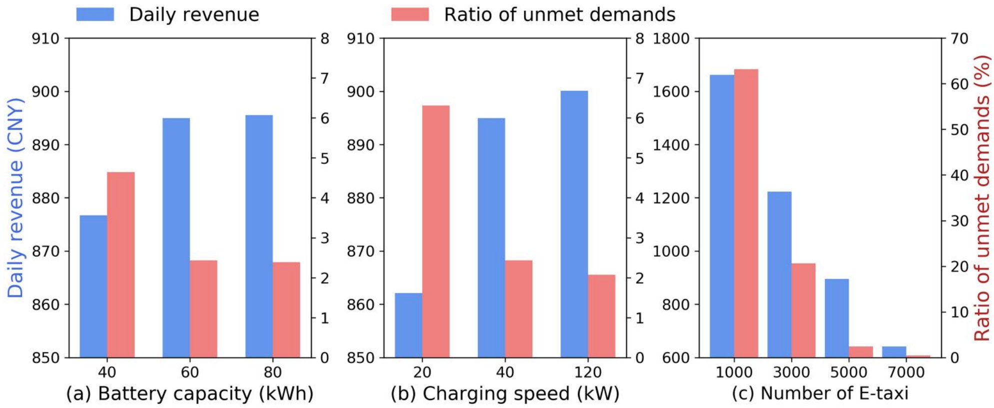

Figure 12. The effect of (a) battery capacity; (b) charging speed; (c) taxi number on daily net revenue and the ratio of unmet taxi demands.

5.5.2. Impact of charging speed

We hypothesized three types of charging infrastructures: low-speed charging (20 kW), high-speed charging (40 kW), and hyper-speed charging (120 kW) to test the sensitivity of the E-taxi system to charging speed, with results shown in Figure 12(b). Generally, faster charging requires less time to complete, thus permitting E-taxis more time on the road. When the charging speed changes from 20 kW to 40 kW, the average daily net revenue is significantly improved from 862 CNY (around 124 US dollar) to 895 CNY (around 129 US dollar). While the unmet taxi demand decreases from 6.3% to 2.4%. When the charging speed further increases to 120 kW, the E-taxis will obtain the highest average daily net revenue of 900 CNY (around 130 US dollar). But the benefit of the increased revenue tends to vanish when the battery capacity and charging speed are further improved as the taxi demand for the whole city is finite.

5.5.3. Impact of the E-taxi fleet size

The number of E-taxis also affects the performance of the system. Figure 12(c) shows the daily net revenue and the percentage of unmet taxi demand with varying E-taxi fleet sizes. As the number of E-taxis increases from 1000 to 7000, the taxi service on the road is significantly improved. The total ratio of unmet demand dropped from 63% to 0.4%. However, the average revenue also decreases from 1662 CNY (around 240 US dollar) to 641 CNY (around 92 US dollar), largely due to more intense competition.

6. Conclusions

An effective and efficient real-time recommendation system is proposed for operating a fleet of centrally managed E-taxis, which accounts for the diverse actions including picking up passengers and serving orders, cruising on the road, waiting in place, and recharging at stations. Spatiotemporal patterns of taxi demand are learned from the real-world trip data of taxis by the spatiotemporal NN method. The ST-DDQN framework is established by coupling the NN model and the double-Q DRL. An online prototype of the recommendation system is developed to implement efficient real-time decision recommendation. An experiment in Shenzhen, China demonstrates the effectiveness and efficiency of the proposed system. The average daily net revenue of E-taxis is significantly higher than those with the baseline methods and in the ICE scenario. The proposed method can not only increase the daily net revenue in the short term but also improve the viability of E-taxis in the long run. This method is suitable for centrally controlled self-driving electric vehicles and can also provide decision support for E-taxi drivers. On the other hand, by integrating a travel information system, real-time traffic information may be considered to further improve the operation of E-taxi fleets. This study also contributes to the literature on spatiotemporal planning regarding the performance optimization of urban transportation systems with reinforcement learning and geospatial big data analytics (Xu *et al.* 2023).

There are also a few directions for future research. First, the aggressiveness of drivers is set uniformly across the driver population in the present study. To account for the heterogeneity of driver behavior, the parameter can be drawn from a pre-defined or empirically fitted probability distribution. Also, largely due to data constraints, traffic conditions are considered exogenous and not affected by the changing actions of the agents. A more comprehensive modeling of the traffic dynamics based on external data and/or some reasonable assumptions could yield more realistic outcomes. Additional model specifications and scenarios could be tested including alternative reward designs to put more emphasis on levels of service for passengers and future scenarios with more advanced battery technologies. White-box models with great interpretability and recent development in expandable and interpretable AI (Fuhrman *et al.* 2022) also provide viable options for the optimization of vehicle fleets and could be applied and compared with the RL framework. Finally, agents in the present study are assumed to be completely rational and their behaviors are modeled mostly independently with full cooperation to the central operator. Occasionally impulsive or sub-optimal actions by the drivers can be included and competitions among agents could be reflected via multi-agent reinforcement learning (Liu *et al.* 2022).

Acknowledgments

The authors would like to thank Prof. Bo Huang and the anonymous referees for their insightful comments that greatly helped improve the paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study is supported and funded by the National Natural Science Foundation of China (No. 42071360; No. 4201367); Key Project of Natural Science Foundation of Shenzhen (No. JCYJ 20220818100200001).

Notes on contributors

Wei Tu is currently an Associate Professor at the Department of Urban Informatics, Shenzhen University. His research interests include automatics recognition of human activity-mobility from multi-source urban data, trajectory modeling, and analysis and optimization. He contributed to the conception of the research idea, the supervision of the experiments, and the writing of the manuscript.

Haoyu Ye is a master student at the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. He contributed to the formulation of the model, the experiments, and the writing of the manuscript.

Ke Mai is a Ph.D. student at the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University. He contributed to the research design and methodology.

Meng Zhou is an Assistant Professor at the School of Intelligent Systems Engineering, Sun Yat-sen University. He contributed to the research design, the supervision of the experiments, and the writing of the manuscript.

Jincheng Jiang is an Associate Research Professor at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. He contributed to the research design and methodology.

Tianhong Zhao is a lecturer at the College of Big Data and Internet, Shenzhen Technology University. He contributed to the research design and methodology.

Shengao Yi is a Ph.D. student at the Department of Geography and Urban Studies, Temple University. He has recently graduated with a M.S. in Urban Spatial Analytics from the University of Pennsylvania. He contributed to the experiments and the writing of the manuscript.

Qingquan Li is currently a Professor at Shenzhen University, Guangdong, China. His research interests include dynamic data modeling in GIS, surveying engineering, and intelligent transportation system. He contributed to the resources and supervision of the project.

Data and codes availability statement

The data and codes that support the findings of this study are openly available in Figshare at https://doi.org/10.6084/m9.figshare.22679116.

References

Abduljabbar, R., *et al.*, 2019. Applications of artificial intelligence in transport: an overview. *Sustainability*, 11 (1), 189.

Bogyrbayeva, A., *et al.*, 2022. A reinforcement learning approach for rebalancing electric vehicle sharing systems. *IEEE Transactions on Intelligent Transportation Systems*, 23 (7), 8704–8714.

Castro, P.S., Zhang, D., and Li, S., 2012. Urban traffic modelling and prediction using large scale taxi gps traces. In: *Pervasive Computing: 10th International Conference, Pervasive*, vol. 10, 57–72. Springer.

Chen, X.J., *et al.*, 2021. Urban hotspots detection of taxi stops with local maximum density. *Computers, Environment and Urban Systems*, 89, 101661.

Dean, M.D., *et al.*, 2022. Synergies between repositioning and charging strategies for shared autonomous electric vehicle fleets. *Transportation Research Part D*, 108, 103314.

Folkestad, C.A., *et al.*, 2020. Optimal charging and repositioning of electric vehicles in a free-floating carsharing system. *Computers & Operations Research*, 113, 104771.

Fuhrman, J.D., *et al.*, 2022. A review of explainable and interpretable ai with applications in covid-19 imaging. *Medical Physics*, 49 (1), 1–14.

Gao, Y., Jiang, D., and Xu, Y., 2018. Optimize taxi driving strategies based on reinforcement learning. *International Journal of Geographical Information Science*, 32 (8), 1677–1696.
Gong, L., *et al.*, 2016. Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science*, 43 (2), 103–114.
Hall, D., Cui, H., and Lutsey, N., 2018. *Electric vehicle capitals: Accelerating the global transition to electric drive*. Briefing by International Council on Clean Transportation (ICCT). https://theicct.org/sites/default/files/publications/EV_Capitals_2018_final_20181029.pdf.
Hasselt, H.V., Guez, A., and Silver, D., 2016. Deep reinforcement learning with double q-learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30 (1), 2094–2100.
Holler, J., *et al.*, 2019. Deep reinforcement learning for multi-driver vehicle dispatching and repositioning problem. *2019 IEEE International Conference on Data Mining (ICDM)*, 1090–1095.
Jindal, I., *et al.*, 2018. Optimizing taxi carpool policies via reinforcement learning and spatio-temporal mining. *2018 IEEE International Conference on Big Data (Big Data)*, 1417–1426.
Kempton, W., 2016. Electric vehicles: Driving range. *Nature Energy*, 1 (9), 1–2.
Kumar, D., *et al.*, 2018. Fast and scalable big data trajectory clustering for understanding urban mobility. *IEEE Transactions on Intelligent Transportation Systems*, 19 (11), 3709–3722.
Liu, C., Chen, C.X., and Chen, C., 2022. Meta: A city-wide taxi repositioning framework based on multi-agent reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems*, 23 (8), 13890–13895.
Liu, L., Andris, C., and Ratti, C., 2010. Uncovering cabdrivers' behavior patterns from their digital traces. *Computers, Environment and Urban Systems*, 34 (6), 541–548.
Liu, Y., *et al.*, 2015. Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers*, 105 (3), 512–530.
Meng, X., *et al.*, 2020. Sequential construction planning of electric taxi charging stations considering the development of charging demand. *Journal of Cleaner Production*, 259, 120794–120794.
Mnih, V., *et al.*, 2015. Human-level control through deep reinforcement learning. *Nature*, 518 (7540), 529–533.
Qu, M., *et al.*, 2014. A cost-effective recommender system for taxi drivers. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 45–54.
Shi, J., *et al.*, 2020. Operating electric vehicle fleet for ride-hailing services with reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems*, 21 (11), 4822–4834.
Singh, A., Al-Abbasi, A.O., and Aggarwal, V., 2022. A distributed modelfree algorithm for multi-hop ride-sharing using deep reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems*, 23 (7), 8595–8605.
Stocker, T., 2014. *Climate change 2013: the physical science basis: Working group i contribution to the fifth assessment report of the intergovernmental panel on climate change*. Cambridge: Cambridge University Press.
Tseng, C.M., Chau, S.C.K., and Liu, X., 2019. Improving viability of electric taxis by taxi service strategy optimization: A big data study of New York city. *IEEE Transactions on Intelligent Transportation Systems*, 20 (3), 817–829.
Tu, W., *et al.*, 2016. Optimizing the locations of electric taxi charging stations: A spatial-temporal demand coverage approach. *Transportation Research Part C*, 65, 172–189.
Tu, W., *et al.*, 2021. Real-time route recommendations for e-taxies leveraging gps trajectories. *IEEE Transactions on Industrial Informatics*, 17 (5), 3133–3142.
Tu, W., *et al.*, 2019. Acceptability, energy consumption, and costs of electric vehicle for ride-hailing drivers in beijing. *Applied Energy*, 250, 147–160.
Vazifeh, M.M., *et al.*, 2018. Addressing the minimum fleet problem in on-demand urban mobility. *Nature*, 557 (7706), 534–538.
Veres, M., and Moussa, M., 2020. Deep learning for intelligent transportation systems: A survey of emerging trends. *IEEE Transactions on Intelligent Transportation Systems*, 21 (8), 3152–3168.
Verma, T., *et al.*, 2017. Augmenting decisions of taxi drivers through reinforcement learning for improving revenues. *Proceedings of the International Conference on Automated Planning and Scheduling*, 27, 409–417.

Wang, H., *et al.*, 2020. Good or mediocre? a deep reinforcement learning approach for taxi revenue efficiency optimization. *IEEE Transactions on Network Science and Engineering*, 7 (4), 3018–3027.

Wang, R., *et al.*, 2015. Taxirec: Recommending road clusters to taxi drivers using rankingbased extreme learning machines. *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 1–4.

Welch, T.F., and Widita, A., 2019. Big data in public transportation: a review of sources and methods. *Transport Reviews*, 39 (6), 795–818.

Xiong, H., *et al.*, 2023. Detecting spatiotemporal propagation patterns of traffic congestion from fine-grained vehicle trajectory data. *International Journal of Geographical Information Science*, 37 (5), 1157–1179.

Xu, M., *et al.*, 2023. Multi-agent reinforcement learning to unify order-matching and vehicle-repositioning in ride-hailing services. *International Journal of Geographical Information Science*, 37 (2), 380–402.

Xu, Z., *et al.*, 2018. Large-scale order dispatch in on-demand ride-hailing platforms: A learning and planning approach. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 905–913.

Yu, X., and Gao, S., 2022. A batch reinforcement learning approach to vacant taxi routing. *Transportation Research Part C*, 139, 103640–103640.

Zheng, Y., 2015. Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology*, 6 (3), 1–41.

Zheng, Y., *et al.*, 2014. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology*, 5 (3), 1–55.

Zhou, T., *et al.*, 2022. GMIX: Graph-based spatial-temporal multi-agent reinforcement learning for dynamic electric vehicle dispatching system. *Transportation Research Part C*, 144, 103886–103886.