

Improving Viability of Electric Taxis by Taxi Service Strategy Optimization: A Big Data Study of New York City

Chien-Ming Tseng, Sid Chi-Kin Chau and Xue Liu

Abstract—Electrification of transportation is critical for a low-carbon society. In particular, public vehicles (e.g., taxis) provide a crucial opportunity for electrification. Despite the benefits of eco-friendliness and energy efficiency, adoption of electric taxis faces several obstacles, including constrained driving range, long recharging duration, limited charging stations and low gas price, all of which impede taxi drivers' decisions to switch to electric taxis. On the other hand, the popularity of ride-hailing mobile apps facilitates the computerization and optimization of taxi service strategies, which can provide computer-assisted decisions of navigation and roaming for taxi drivers to locate potential customers. This paper examines the viability of electric taxis with the assistance of taxi service strategy optimization, in comparison with conventional taxis with internal combustion engines. A big data study is provided using a large dataset of real-world taxi trips in New York City. Our methodology is to first model the computerized taxi service strategy by Markov Decision Process (MDP), and then obtain the optimized taxi service strategy based on NYC taxi trip dataset. The profitability of electric taxi drivers is studied empirically under various battery capacity and charging conditions. Consequently, we shed light on the solutions that can improve viability of electric taxis.

Index Terms—Electric vehicles, big data study, taxi service strategy optimization

I. INTRODUCTION

Taxis are an important part of public transportation system, offering both flexibility of private vehicles and shareability of public transportation. In many cities around the world, there are usually a large of number of taxis, serving the ad hoc demands of commuters. Notably, taxis consume a large amount of fuel. For example, there are over 13,000 taxis operating in New York City, which totally travel over 1.46 billion kilometers each year¹, and consume over 86 million liters of gasoline. As a result, they emit over 242,900 metric tons of CO₂ per year², which is equivalent to the amount of around 25,650 US households' average annual CO₂ emissions³. A viable path toward a low-carbon sustainable society is to promote electrification of transportation, replacing internal combustion engine (ICE) vehicles by more environment-friendly and energy-efficient electric vehicles (EVs). Electrifi-

cation of private vehicles faces many obstacles, such as cost-effectiveness, availability of home charging infrastructure and users' perception. However, electrification of public vehicles (e.g., buses, taxis) would be subject to fewer concerns, with even a greater potential impact than that of private vehicles. First, public vehicles are used more frequently, whose electrification can effectively reduce greenhouse gas emissions. Second, public vehicles are likely to park in common facilities, facilitating the installation of charging stations. Third, public vehicles generally have shorter life cycles due to frequent usage, and hence, are more ready to be replaced.

Major cities worldwide are introducing plans to phase out conventional ICE public vehicles for electric vehicles. For example, Chinese government has initiated several programs to promote electrification of public vehicles for air pollution mitigation [3]. Electric taxi programs were launched in Shenzhen (in 2010) and Beijing (in 2014) to convert taxis to electric vehicles, along with the installation of sufficient EV parking lots and fast charging points. In these programs, the government also offer subsidies to taxi operators. Singapore government plans to roll out a total of 1,000 electric cars to be supported by 2,000 charging points across the city by 2020.

Nonetheless, unlike buses, taxis are often operated as private businesses. Adoption of electric taxis critically depends on the willingness of taxi drivers to switch to electric taxis from conventional ICE taxis. However, it is not clear whether taxi drivers are willing to do so. Despite the initiatives from the governments, there are notable shortcomings of electric taxis:

- 1) **Constrained Driving Range:** One of the barriers preventing wide adoptions of EVs is a shorter driving range. With increasing battery capacity, the driving range has been extended to more than 200 kilometers in production EVs such as Chevrolet Bolt. Generally, the driving ranges of production EVs are sufficient for daily commutes of personal purposes. However, a longer driving range is normally required by logistic vehicles and taxis (e.g., more than 300 kilometers). The driving range of high-end Tesla (as in 2017) may suffice to meet the required driving distance, but are too costly for practical taxis.
- 2) **Long Recharging Duration:** Recharging the battery of EVs can take considerable time. For example, charging Nissan Leaf with 30 kWh battery capacity can take up to 4 hours using mode 3 charging, or half an hour using fast DC charging (without considering queuing delay). Taxis traveling long distances are likely to take more than an hour for recharging between shifts, which is significantly

S. C.-K. Chau is with the Australian National University. X. Liu is with McGill University. (Email: chi-kin.chau@cl.cam.ac.uk, xueliu@cs.mcgill.ca).

This paper appears in IEEE Transactions on Intelligent Transportation Systems (DOI:10.1109/ITITS.2018.2839265).

¹According to New York City taxi trip dataset in 2013 [1].

²Estimated by assuming 67% of New York Yellow taxis as hybrid vehicles and 33% as ICE vehicles, as in 2016.

³The average annual CO₂ emission for US household is 9.5 metric tons [2].

longer than ICE taxis with faster refilling of gasoline.

- 3) **Limited Charging Stations:** Today, the number of charging stations are few. Also, some of charging stations are reserved for specific models or brands with proprietary connectors. The expansion of charging stations is hampered by electrical infrastructure in certain regions. As a result, electric taxi drivers always need sufficient reserve battery capacity in order to be able to return to certain known charging stations, in case of emergence.
- 4) **Low Gas Price:** Nowadays, the oil price has come down considerably from historic heights. This reduces the incentive to adopt EVs, as the gasoline is relatively affordable, despite cheaper and cleaner electricity sources. Unless carbon tax is introduced to mitigate greenhouse gas emissions, gasoline ICE vehicles are still perceived as cost-effective by the public in general.

These shortcomings are likely to dissuade taxi drivers from adopting electric taxis. Particularly, it is not easy to operate a taxi under the constraints of shorter driving range and limited charging stations, in comparison with conventional taxis. In fact, it has been reported in media that taxi drivers tended to shun electric taxis. Without taxi drivers' participation, it is futile to promote electric taxis. Therefore, it is important to provide a viability analysis of electric taxis. Such an analysis can also be used as a basis to determine proper governmental subsidies for electric taxis to promote their adoptions.

In this paper, we identify that a key problem of adopting electric taxis is the ineffective service strategies practiced by today's taxi drivers. In fact, we show that properly optimized taxi service strategies will not suffer from the shortcomings of electric vehicles. Therefore, there is a need to provide an intelligent recommender system to assist taxi drivers to improve their taxi service strategies, and hence, to increase their willingness to switch to electric taxis. In particular, there is a popular trend of ride-hailing mobile apps, which facilitates the computerization and optimization of taxi service strategies, and provide an opportunity of integrating computer-assisted optimized decisions of roaming and navigation to taxi drivers.

A. Modeling Taxi Service Strategy by MDP

The net revenue of a taxi driver (i.e., the revenue from taxi fares minus energy costs) is determined by his/her service strategy of passenger searching and efficiency of passenger delivery. For example, skilful taxi drivers can identify the popular spots for potential passengers, and deliver passengers efficiently by choosing faster routes. Note that the service strategies of taxi drivers can be effectively optimized by utilizing a large historical taxi trip dataset for demand prediction.

To optimize taxi service strategies for electric (or ICE) taxis, we first model computerized taxi service strategy by *Markov Decision Process* (MDP). MDP is a general framework for optimizing sequential decision process in the presence of uncertainty. In summary, we denote a Markov state as the time and location (and possibly battery state) of a taxi, and an action as the driver's decision to travel to the next location (and possibly recharging operations). At each location, there is a probabilistic transition to another location. The transition

is determined by a random event of passenger pick-up. The uncertainty in taxi service strategy is the pick-up location and destination of a passenger, which can be estimated by a historical taxi trip dataset.

This MDP model facilitates the optimization of computerized taxi service strategies by providing computer-assisted decisions to taxi drivers. Since human taxi service strategies are inherently inefficient, optimizing computerized taxi service strategies can potentially improve the net revenues of taxi drivers, particularly in presence of constraints of driving range and charging stations. Computerized taxi service strategies are becoming more feasible, because the increasing adoption of ride-hailing mobile apps, which facilitates the integration of computerized taxi service strategies in a recommender system for taxi drivers using real-time data analytics from historical taxi trip dataset. In this paper, we obtain the optimal policy of MDP that maximizes the revenue of a taxi driver based on New York City taxi trip dataset, and study the profitability of electric taxi drivers under various conditions of battery capacity and charging modes.

B. Summary

Our contributions in this paper are summarized as follows:

- 1) We formulate an MDP to model computerized electric taxi service strategies, with explicit consideration of constraints of EVs, such as battery capacity and locations of charging stations.
- 2) We obtain the optimal policy of the MDP based on a big data study using a large dataset of real-world taxi trips in New York City.
- 3) We study the impact of factors such as battery capacity and charging modes, and locations of charging stations on the net revenues of electric taxi drivers.
- 4) We project our study to understand the benefits of a wider adoption of electric taxis (up to 1000 taxis).

II. BACKGROUND

A. Related Work

Analyzing taxi trip dataset has been considered by several research papers in the subjects of knowledge discovery and cloud-based intelligent transportation systems [4]. One of the popular topics is the profit/revenue improvement for taxi drivers by developing a recommender system for assisting the drivers to find passengers more efficiently. The basic idea is to identify the good taxi service strategies. Several characteristics of taxi service strategies are reported in [5]. Their study shows that searching passengers near the drop-off location of previous passengers results in a higher revenue. They also found that better taxi drivers can deliver the passengers efficiently by choosing a uncongested route. Furthermore, GPS mobility trace from taxis can be used to predict future traffic conditions and optimize the route selections [6]. Also, community detection has been applied to the mobility trace to reveal potential similar passengers' travel patterns, as for social recommendation [7] and improving transportation services [8].

Other studies focus on the specific methods for improving the profit/revenue of the taxi drivers. One approach in [9]

shows that experienced taxi drivers usually wait for passengers at specific locations, and they are usually aware of particular events like train arrivals or ending times of movies.

Instead of recommending separate pick-up locations, a better approach is to maximize the revenue by selecting a route of a sequence of likely pick-up locations at different times. The top-k profitable driving routes can be computed based on a route network with revenues and pick-up probabilities from historical taxi trip data in [10]. To select an optimal route with appropriate actions, Markov Decision Process (MDP) is used to maximize the associated revenue in [11]. The optimal policy of MDP is determined to improve the taxi driver's service strategy. The method of MDP is significantly extended in this paper to consider the constraints of EVs, such as battery capacity and locations of charging stations. Our preliminary study [12] uses a simplified model, whereas this paper presents a more realistic model and a more extensive analysis.

For EVs, limited driving range is a barrier preventing wide adoption. Therefore, the estimation of driving range for EVs has been studied in a number of research papers. The driving range of EVs is highly affected by driving speed and motor efficiency. A black-box model is widely used in the literature to predict the energy consumption of EVs and plug-in hybrid EVs (PHEVs) [13], [14]. Such a black-box model is used in this paper to estimate the energy consumption of electric taxis.

There are other studies that investigated the viability of deploying electric taxis. For example, the return on investment (ROI) for taxi companies transitioning to EVs was studied in [15], which considers the mobility trace of yellow cabs in San Francisco. The prior studies usually assumed that electric taxi drivers will adopt the same service strategies as driving a conventional ICE taxi. On the contrary, our study allows distinctive optimized service strategies for electric taxi drivers, taking into account that EVs have different operating constraints than conventional ICE vehicles.

B. New York City Taxi Trip Dataset

We describe the taxi trip dataset of New York City (NYC) of 2013 that is used in our study. In the following, we list the attributes of dataset that are used in our study. For each data record (i.e., a trip), it is composed of following attributes:

- Taxi ID (also known as medallion ID)
- Trip distance and duration
- Times of pick-ups and drop-offs of passengers
- GPS locations of pick-ups and drop-offs of passengers

We summarize the information of taxi trip dataset in Table I.

TABLE I: New York City taxi trip dataset in 2013.

Attribute	Quantity
Num. of medallions (i.e., rights to operate a taxi)	13437
Annual average traveled distance per taxi	112,600 km
Total num. of trips	175M
Average num. of trips per day	450,000
Average trip distance	4.2 km

The numbers of taxi trips of NYC dataset on different days of 2013 are depicted in Fig. 1a. There are about 450K trips per day and the average trip distance is around 4.2 km. Fig. 1b

displays the pick-up locations on January 16 at 8-9 AM. The k-means algorithm is employed to cluster the pick-up locations by 200 clusters. The sizes of circles indicate the number of pick-up locations. We observe most of pick-ups occur in Midtown Manhattan. Finally, Fig. 1c displays the locations of charging stations in NYC [16] that potentially recharge electric taxis.

III. MARKOV DECISION PROCESS MODEL

In this work, we extend the Markov Decision Process (MDP) framework in [11] to model the computerized service strategy of an electric taxi. MDP facilitates the formulation of computerized taxi service strategies, which can be implemented in a recommender system for taxi drivers. In general, a MDP comprises of a set of states and a set of possible actions at each state. Each action transfers the current state to a new state with a probability and a reward. The objective is to find the optimal actions in the corresponding states that maximize the expected total reward.

A. States and Actions

First, we explain the states and actions of the MDP in our setting. A state for an electric taxi is described by three parameters: current time, current location and battery state, as explained as follows.

- *Current Time*: We consider discrete timeslots. One minute is used as the interval of a timeslot.
- *Current Location*: We consider the locations represented by the nearest junctions, instead of the absolute locations. A road network is constructed using OpenStreetMap (OSM) junction data. Each pick-up or drop-off location is assigned to the nearest junction in OSM. Let \mathcal{N} be the set of all junctions.
- *Battery State*: We consider discrete levels of state-of-charge of battery of the electric taxi. The feasible battery state should be within the range $[\underline{B}, \overline{B}]$.

We denote the location of a taxi at time t by $S(t)$, and the battery state by $B(t)$.

The allowable actions at the current junction are the neighbors of the junction in the road network, and the recharging duration, if the electric taxi is subject to recharging at this junction. We denote an action from junction i to junction j with recharging duration τ at i by $A = (i \rightarrow j, \tau)$, where i and j are neighbors in the road network.

B. State Transition and Objective Function

The basic idea of the MDP for computerized taxi service strategy is illustrated in Fig. 2. Assuming the current location is i , action $A = (i \rightarrow j, \tau)$ is taken. The next location will be j after recharging for a duration τ at i . When entering junction j , there is a probability of not picking up any passenger, after which the taxi driver will make another action. On the other hand, there is a probability of picking up a passenger, with a random destination. The taxi driver will decide if the current battery state is sufficient to deliver the passenger to

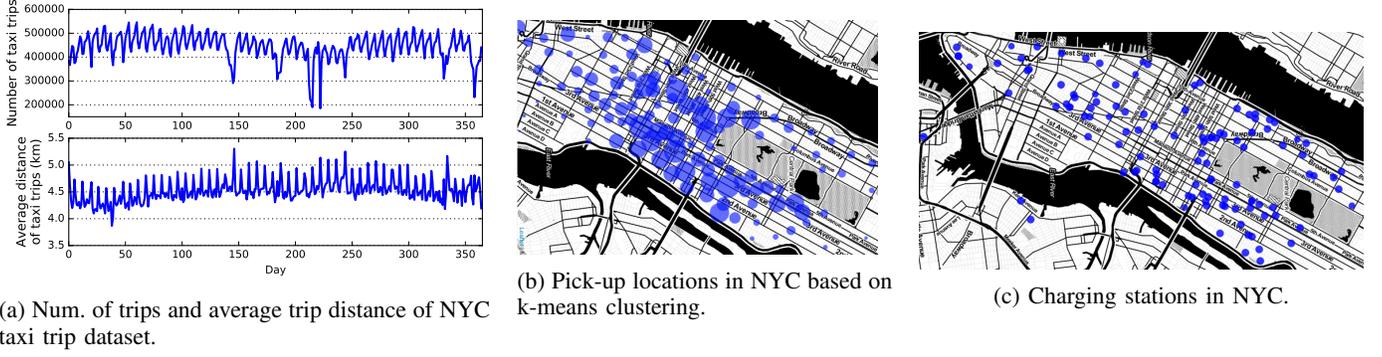


Fig. 1: Overview of NYC taxi trip dataset and locations of charging stations.

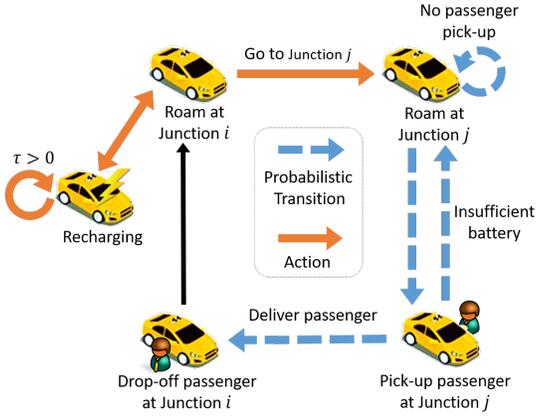


Fig. 2: An illustration of MDP for the computerized service strategy of an electric taxi.

the respective destination, or the trip is discarded. The detailed descriptions of MDP are provided in the following.

First, we define several parameters for the MDP as follows.

- $P_t^p(i)$: The probability of successfully picking up a passenger at junction i at time t .
- $P_t^d(i, j)$: The probability of a passenger commuting from junction i to junction j at time t .
- $T_t^a(A)$: The required time (mins) for executing action A .
- $T_t^t(i, j)$: The required traveling time (mins) from junction i to junction j at time t .
- $E_t^e(i, j)$: The required energy consumption (kW) from junction i to junction j at time t .
- $F_t(i, j)$: The net revenue of transporting passengers from junction i to junction j , which is calculated based on the fare rule of New York taxi and the respective energy costs. There are various surcharges in different times and days, and hence, the net revenue is time-dependent.
- $U_t^a(i, j)$: The energy cost from junction i to j at time t .

Note that some of these parameters (e.g., $P_t^p(i)$, $P_t^d(i, j)$, $T_t^t(i, j)$, $E_t^e(i, j)$) can be estimated from the taxi trip dataset, which will be discussed in the subsequent section.

Next, we formulate a recurrent equation for describing the MDP, namely, Eqn. (1) (as illustrated in Fig. 3).

If the current location is $S(t) = i$, after action $A = (i \rightarrow j, \tau)$ has been taken, the next location will be $S(t') = j$,

where $t' \triangleq t + T_t^a(A)$. The required time of the action $T_t^a(A)$ is computed as follows:

- 1) If recharging duration $\tau = 0$, the taxi directly goes to junction j . The required time of action is given by

$$T_t^a(A) = T_t^t(i, j)$$

- 2) If recharging duration $\tau > 0$, before driving to junction j , the taxi first goes to the nearest charging station $r(i)$ to recharge the electric taxi. The required traveling time is $T_t^t(i, r(i))$ to travel to charging station $r(i)$. Then the electric taxi is recharged for τ duration and next goes from charging station $r(i)$ to junction j , whose required traveling time is $T_{t+T_t^t(i, r(i))+\tau}^t(r(i), j)$. Thus, the total required time of action is given by

$$T_t^a(A) = T_t^t(i, r(i)) + \tau + T_{t+T_t^t(i, r(i))+\tau}^t(r(i), j)$$

Note that if the state-of-charge of battery is insufficient, certain actions are infeasible (e.g., driving to a distant location to pick up passengers). Therefore, an action needs to consider the required energy consumption that can be supported by the current battery state. If the current battery state is $B(t) = b$, after action $A = (i \rightarrow j, \tau)$ has been taken, the new battery state at j will be $B(t') = b' \triangleq \min\{b + \tau C - E_t^e(i, j) - E_t^e(i, r(i)), \bar{B}\}$, where C is the charging rate, and $t' \triangleq t + T_t^a(A)$.

At junction j , there are three possible state transitions:

- (C1) The taxi successfully picks up a passenger at junction j (say, with destination k) and $B(t')$ is sufficient to deliver the passenger to junction k and then to the nearest charging station $r(k)$, if necessary. For each k , the probability is $P_t^p(j)P_t^d(j, k)$, subject to the constraint $E_t^e(j, k) + E_t^e(k, r(k)) + \bar{B} \leq b'$, such that the resultant battery state is always larger than the minimal \bar{B} . Hence, denote the probability of picking up a passenger by probability $P_t^p(j)P_t^s(j)$, where $P_t^s(j)$ is the probability that the destination of passenger is reachable for the taxi under battery constraint, and is computed by

$$P_t^s(j) = \sum_{k \in \mathcal{N}: E_t^e(j, k) + E_t^e(k, r(k)) + \bar{B} \leq b'} P_t^d(j, k)$$

- (C2) The taxi successfully picks up a passenger at junction j , but $B(t')$ is insufficient to deliver the passenger to

$$\begin{aligned}
R^*[t, i, b, A] = & \left(1 - P_{t'}^p(j) + \sum_{k \in \mathcal{N}: E_t^e(j, k) + E_t^e(k, r(k)) + \underline{B} > b'} P_{t'}^p(j) P_{t'}^d(j, k)\right) \cdot R^*[t', j, b'] \\
& + \sum_{k \in \mathcal{N}: E_t^e(j, k) + E_t^e(k, r(k)) + \underline{B} \leq b'} P_{t'}^p(j) P_{t'}^d(j, k) \cdot \left(F_{t'}(j, k) + R^*[t' + T_{j, k, t'}^t, k, b'] - E_{t'}^e(j, k)\right) - U_t^a(i, j) \quad (1)
\end{aligned}$$

junction k and then to the nearest charging station $r(k)$.
The total probability of such a case is

$$\sum_{k \in \mathcal{N}: E_t^e(j, k) + E_t^e(k, r(k)) + \underline{B} > b'} P_{t'}^p(j) P_{t'}^d(j, k)$$

(C3) The taxi cannot successfully pick up a passenger at junction j . The probability is $1 - P_{t'}^p(j)$.

Note that the probability that the taxi does not deliver any passenger (including (C2) and (C3)) is $1 - P_{t'}^p(j) \cdot P_{t'}^s(j)$. The complement of $P_{t'}^s(j)$, i.e., $1 - P_{t'}^s(j)$ is given by

$$1 - P_{t'}^s(j) \triangleq \sum_{k \in \mathcal{N}: E_t^e(j, k) + E_t^e(k, r(k)) + \underline{B} > b'} P_{t'}^d(j, k)$$

Hence, we obtain

$$\begin{aligned}
& 1 - P_{t'}^p(j) \cdot P_{t'}^s(j) \\
= & 1 - P_{t'}^p(j) + P_{t'}^p(j) \cdot (1 - P_{t'}^s(j)) \\
= & 1 - P_{t'}^p(j) + \sum_{k \in \mathcal{N}: E_t^e(j, k) + E_t^e(k, r(k)) + \underline{B} > b'} P_{t'}^p(j) P_{t'}^d(j, k)
\end{aligned}$$

For (C1), the taxi driver will receive a fare of amount $F_{t'}(j, k)$, and the next location of the taxi becomes $S(t' + T_{j, k, t'}^t) = k$. For (C2) and (C3), the taxi driver will not receive any fare, and will decide to drive to another location or possibly recharge the taxi.

The objective of the MDP is to maximize the total expected net revenue. Note that the net revenue of the action is the received fare minus the energy cost of the action. The expected net revenue for an action $A = (i \rightarrow j, \tau)$ at state (t, i, b) is denoted by $R^*[t, i, b, A]$, which can be computed recurrently in Eqn. (1), where

- $t' = t + T_t^a(A)$ and $b' = \min\{b + \tau C, \bar{B}\}$.
- $R^*[t, j, b'] = \max_A R^*[t, j, b', A]$ is the maximal expected net revenue in state (t, j, b') over all possible actions.
- $U_t^a(i, j)$ is the energy cost, as computed as follows:

- 1) If recharging duration $\tau = 0$, the taxi directly goes to junction j . The energy cost is $U_t^a(i, j) = E_t^e(i, j) \cdot U$, where U is the unit price, such that $U = 20$ cent/kWh for electricity and $U = 2.5$ USD\$/gallon for gasoline.
- 2) If recharging duration $\tau > 0$, the taxi goes to the nearest charging station $r(i)$ to recharge the electric taxi at charging rate C . The energy cost of the action is given by

$$U_t^a(i, j) = (E_t^e(i, r(i)) + E_{t+T_{i, r(i)}^t}^e(r(i), k) + \tau \cdot C) \cdot U$$

We seek to devise an optimal policy π for the MDP that maximizes the expected net revenue:

$$\pi(t, S(t), B(t)) = \arg \max_A R^*[t, S(t), B(t), A] \quad (2)$$

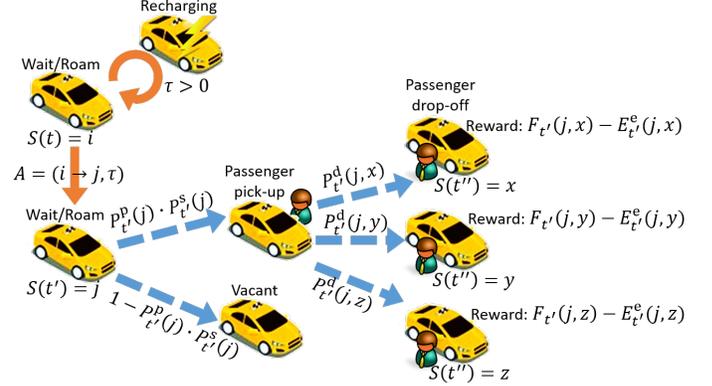


Fig. 3: An illustration for the recurrent equation Eqn. (1).

To obtain the optimal policy for the MDP, one can use dynamic programming. The dynamic programming algorithm starts from the last timeslot and then works backwards to the beginning timeslot. For example, to solve the optimal policy for a morning shift, the algorithm starts to solve the maximal expected net revenue at the end of shift, and works backwards.

IV. MARKOV DECISION PROCESS PARAMETERS

In this section, we estimate several parameters of MDP (e.g., $P_t^p(i)$, $P_t^d(i, j)$, $T_t^t(i, j)$, $E_t^e(i, j)$) from NYC taxi trip dataset.

A. Driving Speed Network

First, we construct a driving speed network from the NYC taxi trip dataset, for the following purposes:

- 1) To estimate the traveling time from each junction to the nearest charging station.
- 2) To estimate the energy consumption of a taxi for a trip.

Note that traveling time and driving speed are time-dependent parameters, since they are highly affected by traffic condition, which is estimated from historical trip data. For example, the traveling time between the same pair of junction i and junction j will be higher in office hours and much lower at midnight.

The first step of constructing the driving speed network is to determine the driving path of a taxi. Spatialite [17] is used to calculate the shortest path for each pair of pick-up and drop-off locations. Spatialite utilizes OpenStreetMap (OSM) data. A resulting path comprises a list of edges (i.e., segments) described by two junctions. We then compare the recorded trip distance in the taxi trip dataset to the computed shortest path distance. If the difference is greater than 300 meters, the record is discarded since the driver is likely to take other route. For each computed path, the segments of a path are labeled with the average speed using recorded traveling time

and distance. We can obtain the average speed for each taxi trip record. Each segment has several average speeds by different trips. We select the highest speed to represent the driving speed of the segment, since this is usually the speed with minimal obstacles. Driving speed networks at different times are visualized in Fig. 4. We observe there is relatively more congested traffic in 9 to 10 AM or 4 to 5 PM.

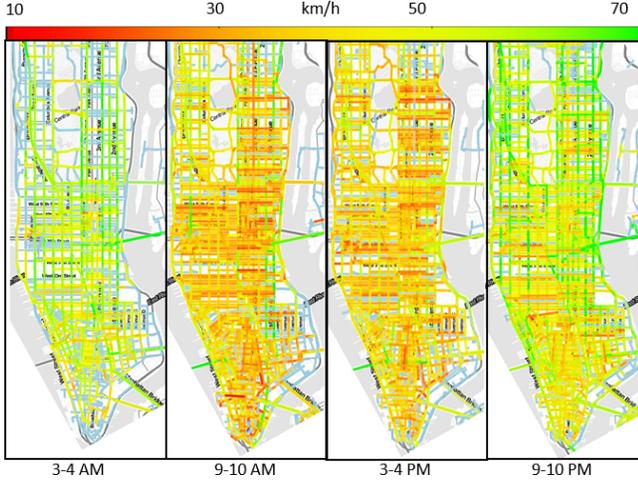
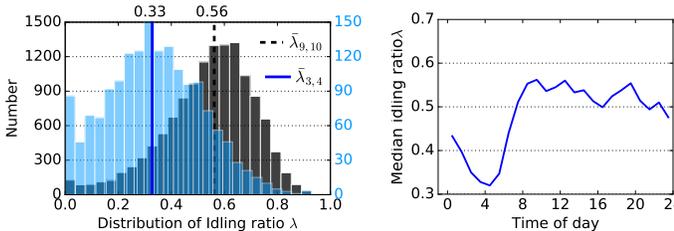


Fig. 4: Visualizations of driving speed networks.

Given the driving speed network, we can estimate the driving time from the network. We can also estimate the idling time of each trip by subtracting the estimated driving time from the recorded traveling time. The detailed steps for calculating the idling time are described as follows:

- 1) Average traveling time $T_t^t(i, j)$: There may be several trips start from junction i to junction j . However, their traveling times may be slightly different. We average the traveling time of these trips.
- 2) Driving time $T_t^d(i, j)$: The shortest path from junction i to junction j is determined by Spatialite. Then, the driving time in each segment is computed by its distance and the driving speed from the driving speed network.
- 3) Idling time $T_t^i(i, j)$: The idling time of a trip is obtained by subtracting the driving time from the average traveling time, $T_t^i(i, j) = T_t^t(i, j) - T_t^d(i, j)$



(a) Distribution of idling ratios for 3-4 AM and 9-10 AM (b) Median idling ratios over a day.

Fig. 5: Hourly distribution of idling ratios and median idling ratio over a day.

To understand traffic conditions, define a metric called the idling ratio of each source and destination pair by $\lambda \triangleq \frac{T_t^i(i, j)}{T_t^t(i, j)}$.

Denote by $\bar{\lambda}_{t_1, t_2}$ the median of idling ratio between time t_1 and t_2 in the distribution. Fig. 5 shows the distribution of idling ratios. We observe that the median is 56% for 9-10 AM, but only 33% for 3-4 AM due to less traffic.

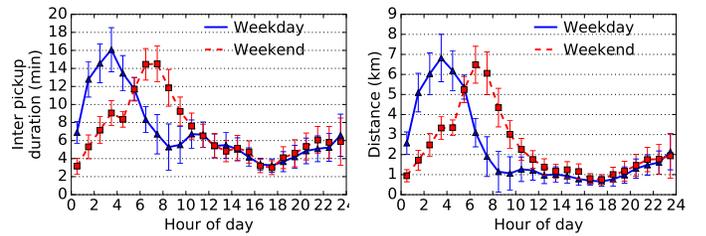
B. Passenger Pick-up Probability $P_t^p(i)$

The passenger pick-up probability describes the chance of a taxi driver can pick up a passenger at junction i at time t . Following the idea in [11], we use the numbers of taxis and pick-ups around a particular junction to calculate the pick-up probability $P_t^p(i)$ in τ mins. First, denote the number of pick-ups at junction i from time t to $t+\tau$ by $N_{t:t+\tau}^p(i)$. To estimate the number of taxis around junction i in τ mins, denote the number of drop-offs from time $t-\tau$ to $t+\tau$ within δ kilometers distance from junction i by $N_{t-\tau:t+\tau}^d(i)$. Assuming the taxis are vacant after dropping off the passengers and are roaming immediately around junction i within δ kilometers in τ mins. Thus, pick-up probability $P_t^p(i)$ can be estimated by

$$P_t^p(i) = \frac{N_{t:t+\tau}^p(i)}{N_{t:t+\tau}^p(i) + N_{t-\tau:t+\tau}^d(i)} \quad (3)$$

The suitable parameters τ and δ can be obtained from the historical taxi trip dataset. For example, τ can be estimated by the average inter-pick-up duration, the time interval between consecutive pick-ups of a taxi. Using the average driving speed, δ can be estimated by the reachable distance in the average inter-pick-up duration. Fig. 6a depicts the average inter-pick-up durations for weekdays and weekends. We observe that it takes more time to find a passenger at 4 AM on weekday and at 7 AM at weekends. Fig. 6b depicts the respective reachable distance in inter-pick-up duration.

In the following study, we set time-varying τ and δ according to the average inter-pick-up duration and the respective reachable distance from taxi trip dataset for each hour.



(a) Inter-pick-up durations for weekdays and weekends. (b) Reachable distances in inter-pick-up durations.

Fig. 6: Parameters for estimating pick-up probability.

C. Passenger Destination Probability $P_t^d(i, j)$

The passenger destination probability describes the chance that a passenger needs to commute from one junction to another junction. This probability is time-dependent, because, for example, passengers are more likely to commute from living places to offices in working hours. One-hour timeslot is used to estimate passenger destination probability from taxi trip dataset. In each timeslot, we obtain the number of trips between each pair of source and destination, and

then is normalized by the total number of trips. Denote the destination probability from junction i to junction j at time t by $P_t^d(i, j)$. Denote the number of pick-ups at junction i by $N_t^p(i)$, and the number of corresponding drop-offs at junction j by $N_t^d(i, j)$. The passenger destination probability from junction i to junction x is estimated by

$$P_t^d(i, j) = \frac{N_t^d(i, j)}{N_t^p(i)} \quad (4)$$

D. Energy Consumption $E_t^e(i, j)$

We use a black-box approach to estimate the energy consumption for EVs, based on the work in [13], [14]. The energy consumption model is based on the average driving speed and auxiliary loading. The total energy consumption can be decomposed into moving energy consumption and auxiliary loading energy consumption, which can be estimated by multivariate linear models (see [13], [14] for details):

$$E_t^e(i, j) = E_t^{mv}(i, j) + E_t^{ax}(i, j) \quad (5)$$

$$E_t^{mv}(i, j) = \beta(\alpha_1 v_t(i, j)^2 + \alpha_2 v_t(i, j) + \alpha_3) \cdot D(i, j) \quad (6)$$

$$E_t^{ax}(i, j) = \ell_t T_t^x(i, j)/60 \quad (7)$$

where $v_t(i, j)$ is the driving speed between junctions i and j at time t , obtained from driving speed network. $D(i, j)$ is the driving distance between junction i and junction j .

The auxiliary loading ℓ_t is highly affected by weather temperatures which is time variant. The auxiliary loading can be estimated from the historic weather temperature and the average auxiliary loading measurements at particular temperatures⁴. According to New York historical weather and suggested power load, the average auxiliary loading is between 1.5 to 1 kW. The parameter β represents aggressiveness factor to capture the driving behavior. Driver behavior has an impact on the energy consumption of vehicles, as driving range will be significantly decreased by aggressive acceleration and deceleration. Mild driving behavior can save up to 30% to 40% energy consumption comparing with aggressive driving behavior [19], [20]. Therefore, we define three classes of driving behaviors: i) mild drivers ($\beta = 0.8$), ii) normal drivers ($\beta = 1$), and iii) aggressive drivers ($\beta = 1.2$). Based on previous work [13], the parameters of energy consumption model for Nissan Leaf are set as $\alpha_1 = 0.1554$, $\alpha_2 = -5.4634$, $\alpha_3 = 189.297$.

E. Energy Consumption $E_t^e(i, r(i))$

The electric taxis should arrive at each junction with certain battery state, which can guarantee them to reach the nearest charging stations. The locations of NYC charging station data are obtained from [16]. We consider the charging stations for general EVs. Note that there are other charging stations requiring memberships, and are not considered in this study. To estimate the minimum required energy consumption $E_t^e(i, r(i))$ to the nearest charging station $r(i)$ at junction i at time t , the minimum distance between the junction and the nearest charging station is obtained as follows:

- 1) Spatialite is used to find the nearest charging station $r(i)$ for junction i in the road network by the shortest distance.
- 2) The shortest distance is converted into the required driving time based on the driving speed network.
- 3) The median idling ratio $\bar{\lambda}$ is used to estimate the idling time at time t .
- 4) Given the driving speed network and idling time, the energy consumption $E_t^e(i, r(i))$ is obtained by Eqn. (5).

F. Taxi Net Revenue $F_t(i, j)$

The fares are calculated according to the rules for New York taxis. Since there are different kinds of surcharge based on times and days, the fare is time-dependent, because of various surcharges⁵. The net revenue of a trip can be calculated by deducting fuel/electricity cost from the revenue. Therefore, the net revenue of a trip from junction i to junction j at time t is

$$F_t(i, j) = F_t^R(i, j) - E_t^e(i, j) \cdot U \quad (8)$$

where $F_t^R(i, j)$ is the recorded amount of base fare plus the surcharges from i to j at time t , and U is the unit price.

G. Charging Rate C

Two types of charging rates are considered in this study: mode 3 charging and (direct current) fast charging. Currently, mode 3 charging is more common than fast charging. The charging power of mode 3 charging is 6.6 kW (e.g., for Nissan Leaf), whereas the charge power of fast charging is 50 kW.

V. EVALUATION BASED ON NYC TAXI TRIP DATASET

In this section, we apply the MDP to optimize computerized taxi service strategies and evaluate the improvement in net revenues using NYC taxi trip dataset. We first examine the net revenue of conventional ICE taxis and improvement by MDP under a basic setting with complete knowledge of taxi trip information for one single taxi, which represents the best-case scenario. Next, we study a similar setting for electric taxis. Then, we relax the basic setting by more realistic settings: (1) using only historical data as training dataset, (2) an extension to multiple taxis, and (3) considering different driving behavior.

A. Basic Setting of ICE Taxi

Setting: This section presents an evaluation study based on one-day data of January 9 2013 in the NYC taxi trip dataset. In Sec. V-G, an evaluation using a whole year's data will be presented. First, we note that the NYC taxi trip dataset has only records of trip distance and duration, and pick-up and drop-off information. There is no full mobility data trace of taxis, in particular when the taxis are roaming without passengers. It

⁵The initial charge is \$2.50. Plus 50 cents per 1/5 mile or 50 cents per 60 seconds in slow traffic or when the taxi is stopped. 50-cent MTA State Surcharge is required for all trips that end in New York City. Another 30-cent Improvement Surcharge is required. Daily 50-cent surcharge is required from 8pm to 6am. \$1 surcharge is required from 4pm to 8pm on weekdays, excluding holidays. Toll fees are ignored since the taxi driver will not receive any revenue from tolls.

⁴ See [18] for an empirical measurement study

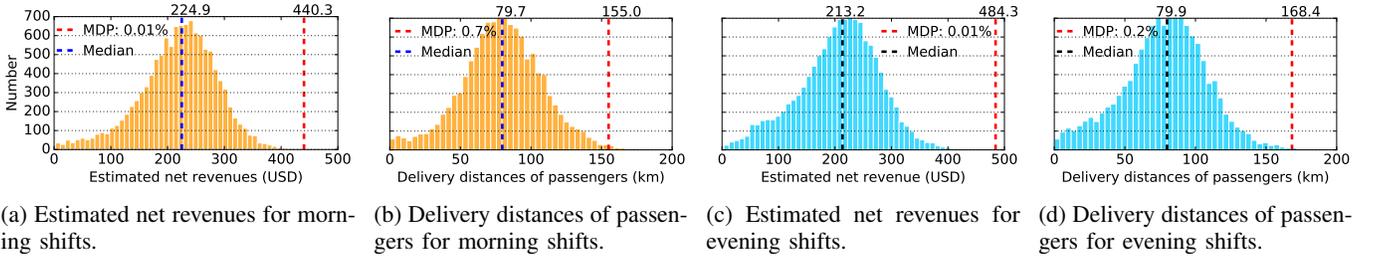


Fig. 7: Distributions of estimated net revenues and delivery distances of passengers.

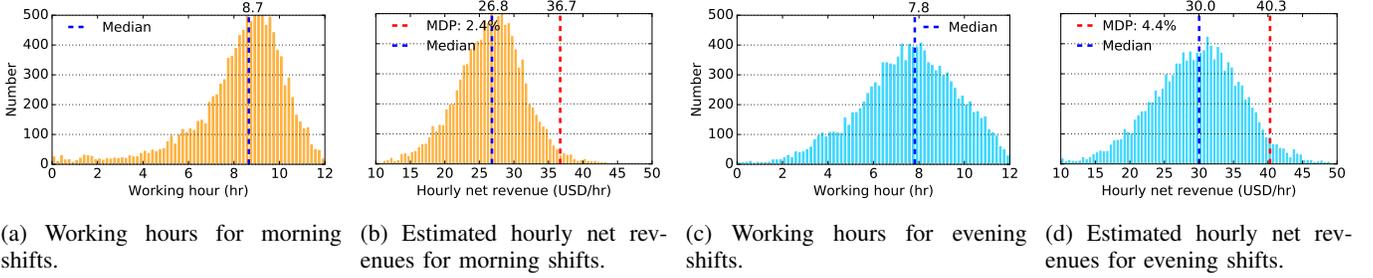


Fig. 8: Distributions of working hours and estimated hourly net revenues.

is difficult to estimate the exact total travel distance (i.e., including roaming and passenger delivery). Hence, we estimate a lower bound for the total travel distance by connecting the shortest path between a drop-off location and a subsequent pick-up location. As such, we obtain an optimistic estimation of net revenue (i.e., revenue minus fuel cost) by the lower bound of total travel distance. We consider a basic setting, such that the optimal policy of MDP is employed in one single taxi, based on complete knowledge of taxi trip information on the same day from the dataset. In Sec. V-D, using historical data for prediction and multiple taxis will be presented.

Note that it is challenging to evaluate the exact performance of modified taxi behavior using historical dataset. For example, when a passenger is picked up by a taxi with modified behavior, who was originally picked up by another taxi in the dataset, it is not clear how original taxi should behave in the evaluation. Therefore, we consider a simple approach of evaluation, such that other taxis always follow the recorded trajectories as in the dataset, no matter picking up the supposed passengers of the dataset or not. Although this will not attain absolute accuracy, this is a simple approach without the knowledge of the disrupted behavior of other taxis in real life. Note that if we only modify the behavior of a small number of taxis, then this simple approach will give rather accurate evaluation.

Also, refueling is not considered for ICE taxis, because ICE taxi drivers normally fill up the gas tanks between the shifts⁶. Then the MDP model for an ICE taxi is identical to that of an electric taxi in Sec. III, but without recharging decisions.

Observations: Based on the NYC taxi trip dataset, Fig. 7a shows the distribution of (optimistically) estimated net revenues from all the trips (with 11746 taxi drivers) for morning

shifts. The blue dashed line indicates the median of taxi drivers. We observe that 50% drivers earn above USD\$223. The red dashed line indicates the expected estimated net revenues when a taxi driver follows the optimal policy of MDP assuming 12 working hours. This taxi driver is expected to earn USD\$440. Therefore, optimizing the taxi service strategy enables a taxi driver to earn at most among the top 0.01%. Fig. 7b shows the delivery distances of passengers per taxi drivers. More than 50% taxis travel more than 79 kilometers for passenger delivery. By optimizing taxi service strategy, a taxi driver is expected to travel up to 155 kilometers for passenger delivery. Fig. 7c-7d show the distributions for evening shifts. The median net revenue is smaller than that of morning shifts because of shorter working hours (Fig. 8c). Also, we observe that the median delivery distances of passengers for the evening shift is similar to that of morning shifts.

The computation of expected net revenue of the optimal policy of MDP assumes 12 working hours. The distribution of working hours for morning shifts is shown in Fig. 8a. We observe that most of drivers work less than 12 hours, and the median working hours on the day is 8.7 hours. For a normalized comparison, we also study the hourly net revenues, instead of net revenues per shifts. The distribution of estimated hourly net revenues for morning shifts is presented in Fig. 8b. We observe that the hourly net revenue of MDP driver is the top 5% in both shifts. We notice that higher hourly net revenue is due to shorter working hour with long trips. Fig. 8c shows that taxi drivers have shorter working hours for evening shifts, but their hourly net revenues, because of extra surcharge for evening shifts.

Ramifications: Optimizing taxi service strategies can significantly improve the profitability of taxi drivers. Our evaluation based on a basic setting shows that optimized service strategy for a conventional ICE taxi can earn at most among the

⁶Most NYC taxis operate in two shifts per day. Each normally lasts for 12 hours. More than 40% of taxi drivers change shifts at around 5 AM or 5 PM. In this study, we assume that a morning shift is from 5 AM to 5 PM, whereas an evening shift is from 5 PM to 5 AM.

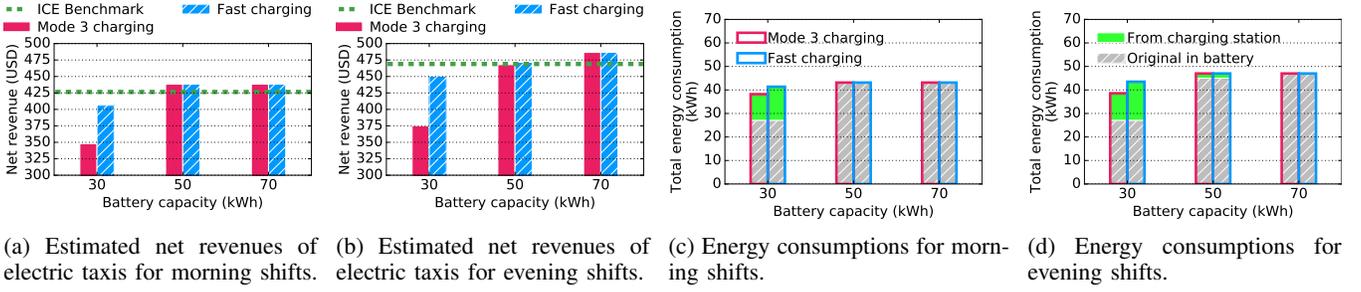


Fig. 9: Estimated net revenues and energy consumptions of electric taxis.

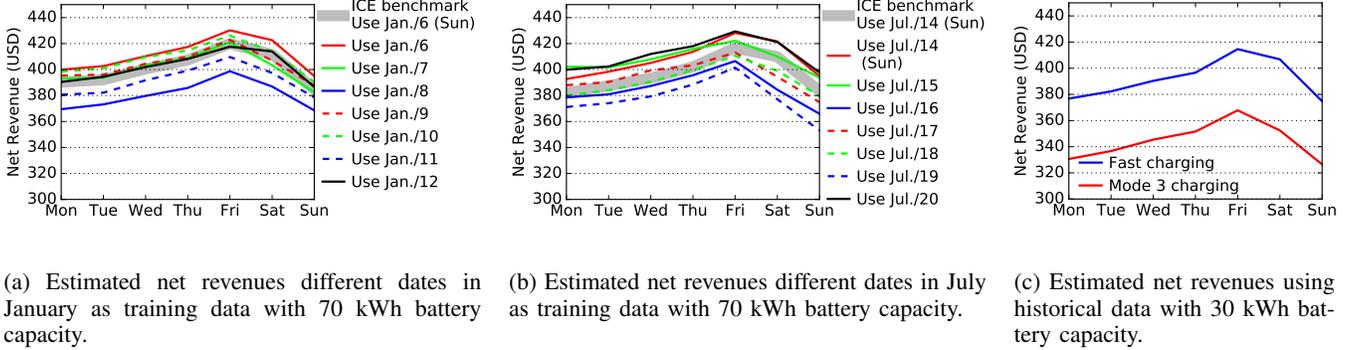


Fig. 10: Estimated net revenues using historical data for electric taxis.

top 0.1%. Although this represents the best-case evaluation, the subsequent sections will relax to more realistic settings, and yet still show a considerable advantage.

B. Basic Setting of Electric Taxi

Setting: In this section, we apply a similar basic evaluation based on the data of January 9 2013 to electric taxis. We employ the energy consumption model of Nissan Leaf [13]. In fact, the most determining factor of performance of EVs is battery capacity. Hence, the energy consumption model of Nissan Leaf suffices to provide a generic estimation of energy consumption of electric taxis. Usually, the EVs will not allowed to be overly re/discharged to protect the battery. Therefore, we set the available battery level from 5% to 95% of the capacity. We consider typical settings of battery capacity for EVs (e.g., 30 kWh, 50 kWh, 70 kWh). Each setting can affect the recharging decisions and net revenues considerably.

Observations: Figs. 9a-9b show the estimated net revenues for electric taxis under different battery capacities. The blue bars represent the net revenues using fast charging, while the red bars represent those using mode 3 charging. We observe that electric taxis equipped with 50 kWh battery can make comparable net revenues with traditional ICE taxis using fast charging for morning shifts. Note that in general smaller batteries require more frequent recharging, which can reduce revenue. EVs with smaller batteries are cheaper. The net revenue gap between using fast charging and mode 3 charging is smaller when the battery capacity increases. The estimated net revenue reaches USD\$438 when battery capacity is above 50 kWh. The net revenue is higher than that of ICE taxis

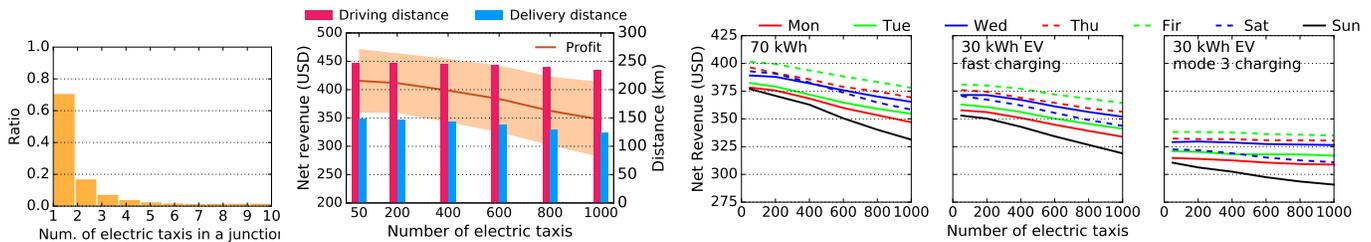
using optimized service strategies (i.e., USD\$426 benchmark), because electricity cost is cheaper.

Figs. 9c-9d show the driving distances and energy consumptions under different battery capacities. The blue bordered bar represent the driving distances using fast charging while red bordered bars represent those using mode 3 charging. The green portions represent the amount of charging energy received from charging stations, while gray portions represent the amount from initial batteries. We observe that the total driving distance is around 242 kilometers without recharging for morning shifts. At night, the electric taxis are expected to drive longer distances because of less traffic. The required energy consumption without charging for morning shifts is 43 kWh, which can be provided by 50 kWh battery (i.e., 45 kWh usable capacity) without recharging. For evening shifts, the required energy consumption increases to 45.1 kWh. Therefore, electric taxis with 50 kWh battery are then required to recharge during shifts.

Ramifications: Optimizing taxi service strategies for electric taxis can improve the profitability of taxi drivers. But the effect depends on the battery capacity. With more capacity (e.g., 50 kWh, 70kWh), the taxi driver can earn comparable net revenue with the one of ICE taxi using optimized service strategy. It is because that recharging will incur inefficiency for electric taxis with a low capacity battery.

C. Using Historical Data for Prediction

Setting: The previous basic evaluation of net revenues is based on MDP using the complete knowledge, which requires knowing the pick-up demands and locations in a-priori manner. However, complete information is difficult to obtain in



(a) Histogram of number of taxis in a junction. (b) Estimated net revenues for multiple electric taxis with 70 kWh battery capacity. (c) Estimated net revenues for multiple electric taxis over year 2013.

Fig. 11: Estimated net revenues for multiple electric taxis.

practice. A more practical approach is to use only historical data as training dataset for MDP, and then obtain an optimal policy as a heuristic for other days. In the following, we use the optimal policy of MDP obtained from 6th January to 12th January (i.e., the first week after 1st January) as training data. Then we employ the policy to all morning shifts in the year in the evaluation.

Observations: Fig. 10 shows the estimated net revenue using one-day training data on different days of a week. We observe that the highest net revenue occurs on Friday while the lowest net revenue occurs on Sunday, because of more passengers on Fridays. The figures also show the benchmark for ICE taxis using historical data (i.e., gray band). In particular, Fig. 10a shows the net revenue of 70 kWh battery capacity using different dates in January as training data. We observe that the training data from 6th January performs the best while training data from 8th January performs the worst. A taxi driver can receive 7.5% higher net revenue using 6th January data.

Fig. 10c shows the net revenues with the 30 kWh battery capacity using training data from 6th January. We observe that 12% higher net revenue can be obtained using fast charging than using mode 3 charging. Electric taxis with 70 kWh battery capacity can obtain 2.8% higher than 30 kWh battery capacity using fast charging.

Ramifications: Using historical data for prediction, instead of complete knowledge, will inevitably reduce the effectiveness. However, this creates a similar effect on ICE taxis that also use historical data. Hence, optimizing taxi service strategies for electric taxis using historical data still achieves comparable net revenues as that of ICE taxis.

D. Multiple Electric Taxis

Setting: The optimal policy from MDP has been previously employed in one single taxi. Next, we employ the optimal policy to multiple taxis. The idea is to allow multiple electric taxis adopt the optimal policy from MDP, while ensuring the number of taxis being sent to each location is constrained. Otherwise, this leads to over-provision of taxis at certain locations. This simple constraint allows us to decouple the individual MDP decisions. Otherwise, considering a large complex problem will be intractable. In practice, we may display the potential net revenue of each junction to the taxi

drivers. The junction will become less desirable, when the number of taxis currently present exceeds a certain threshold. Hence, they would not prefer to go to the junction.

We first empirically study the distribution of number of taxis at all the junctions over time from the dataset. We then set of limit of the number of taxis at each junction according to the mean number of taxis at each junction from the dataset. To satisfy the constraint, some electric taxis would need to follow the second-best decisions in the optimal policy. Each taxi state is initialized by the junction and the time according to the dataset. The state of each taxi is tracked and the passenger pick-up probability $P_t^p(i)$ is recomputed using Eqn. (3). We use the optimal policy based on the data of 6th January.

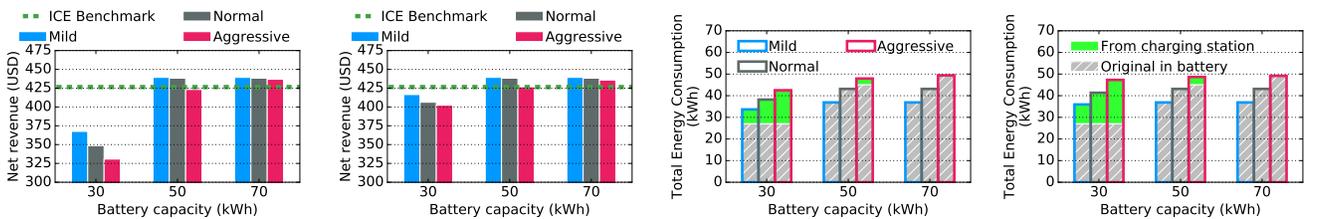
Observations: Fig. 11a displays the histogram of number of taxis in a junction. We observe that the number of taxis in each junction is less than 7 by 99% of time. We set of limit of the number of taxis at each junction according to the mean number of taxis at each junction from the dataset.

Fig. 11b shows the net revenues of different numbers of electric taxis using the optimal policy of MDP on 9 Jan. We observe that the net revenue drops to \$USD 350 when 1000 electric taxis use the optimal policy of MDP. The red bar indicates the total driving distance of the taxis and blue bar indicates the passenger delivery distance. We observe that the delivery distance drops but the total driving distance remains relatively steady. This implies that the increase of roaming distance is due to a lower passenger pick-up probability. Fig. 11c shows the average net revenue of multiple taxis over entire year of 2013. We observe that the highest net revenue occurs on Fridays while the lowest occurs on Sundays. We also observe that the net revenue is less affected by the number of taxis when mode 3 charging is used. This is because that the electric taxis require frequent recharging, which may result in less available taxis, and hence, a higher pick-up probability.

Ramifications: If the optimal policy of MDP is deployed up to 1000 electric taxis, then the net revenues will decrease, as a result of diminishing advantage of computerized service strategies. These 1000 taxi drivers can still earn as top 1.7% among traditional taxi drivers without computerized service strategies.

E. Considering Driving Behavior

Setting: Driving behavior plays an important role in energy consumption of vehicles. Aggressive driving behavior results



(a) Estimated net revenues of different driving behaviors using mode 3 charging.

(b) Estimated net revenues of different driving behaviors using fast charging.

(c) Energy consumption of different driving behaviors using mode 3 charging.

(d) Energy consumption of different driving behaviors using fast charging.

Fig. 12: Estimated net revenues and energy consumptions of different driving behaviors.

in more energy consumption. Furthermore, higher energy consumption rate induces more frequent recharging of EVs, which reduces the net revenues of the taxi drivers. We study three classes of driving behaviors: i) mild drivers ($\beta = 0.8$), ii) normal drivers ($\beta = 1$), and iii) aggressive drivers ($\beta = 1.2$).

Observations: Fig. 12 shows the estimated net revenues of different driving behaviors for morning shifts. Fig. 12a shows the estimated net revenues of different driving behaviors using mode 3 charging. Mild drivers can receive 14% higher net revenue than aggressive drivers when driving 30 kWh Leaf using mode 3 charging. However, the net revenue is less affected by different drivers when the battery capacity is sufficiently large to eliminate recharging during a shift. Fig. 12b shows the estimated net revenues using fast charging. We observe that the net revenue is also less affected by different driving behaviors because of shorter recharging duration. Fig. 12 shows the energy consumption of different driving behaviors. We observe that aggressive drivers consume around 11 kWh more energy than mild drivers.

Ramifications: Although the aggressive drivers consumes 20% more energy which only results in \$USD2.2 difference for morning shifts. The result shows that the driving behavior only has a higher impact on the net revenue when the battery capacity is insufficient to eliminate recharging during a shift.

F. Considering Different Gas Prices

Setting: To complete the study of viability of electric taxis, we provide a study of ICE taxis' net revenue under different gas prices. Note that the current gas price in USA is around USD\$2.5 per gallon, while the current gas price in China is around 7.2 RMB per liter, which is equivalent to \$4.5 USD per gallon. We analyze the outcomes of three different gas prices (i.e., \$2.5 USD/G, \$3.5 USD/G, \$4.5 USD/G) considering the optimal policy of MDP for an ICE taxi.

Observations: Fig. 13 compares the annual net revenues of ICE taxi under different gas prices, with that of electric taxis using different charging options. We observe that the annual net revenue with gas price \$2.5 USD/G (i.e., the leftmost bar) is slightly higher (about USD\$ 4000 higher) than that with gas price \$4.5 USD/G. We also observe that the comparable net revenue can be achieved by 30 kWh EV with fast charging when gas price increases to \$4.5 USD/G. However, the annual net revenue of 30 kWh EV with mode 3 charging is much lower (about 14% lower), even when the gas price is high.

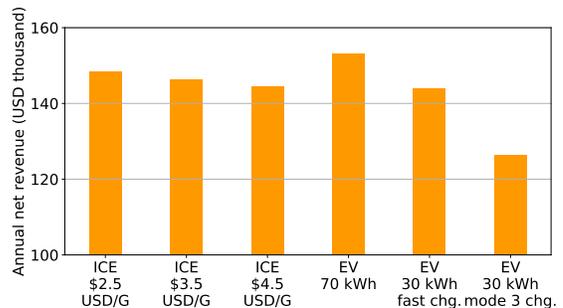


Fig. 13: Annual net revenues under different gas prices.

Ramifications: We observe that when the gas price increases, ICE taxi becomes a less attractive option since its net revenue decreases. The net revenue of 70 kWh EV is around 3% higher than ICE taxi when gas price is \$2.5 USD/G, while it is around 6.6% higher than ICE taxi when gas price increases to \$4.5 USD/G.

G. Annual Evaluation of NYC Taxi Trip Dataset

1) Net Revenue Evaluation:

Setting: We employ the optimal policy from 6th January to different numbers of electric taxis and estimate the annual net revenues. Fig. 15 shows the distribution of annual working hours, we observe that the median annual working hour is around 1800 hours, but many drivers work more than 4300 hours, equivalent to working almost 12 hours a day. Therefore, we consider taxi drivers working every morning shift (i.e., 4380 working hours) to estimate their net revenues.

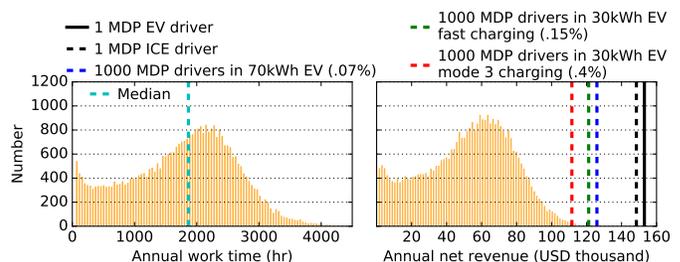


Fig. 15: Annual working hours and estimated net revenues of taxi drivers in 2013.

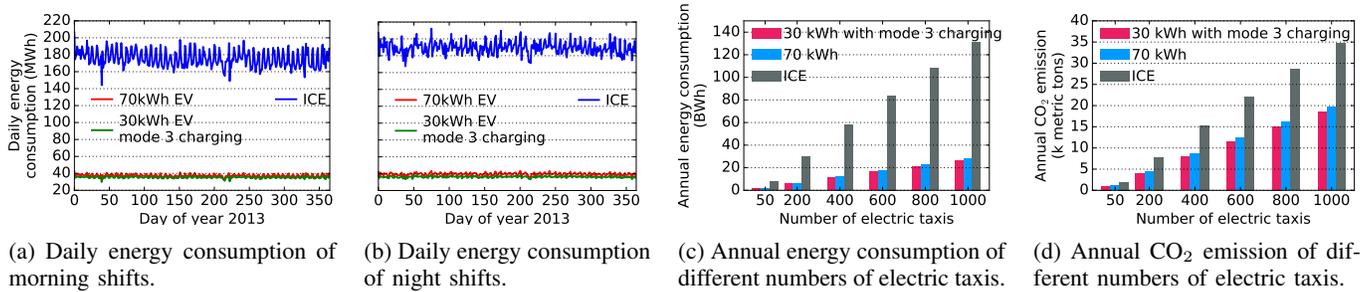


Fig. 14: Daily and annual energy consumption and CO₂ emission.

Observations: The right figure in Fig. 15 shows the estimated net revenues of different taxi drivers using the optimal policy of MDP. There are some observations:

- The case of one electric taxi driver using the optimal policy of MDP can earn 3% higher than that of one ICE taxi driver.
- The average net revenue of case of 1000 electric taxis with 70 kWh battery is ranked top 0.07% among traditional taxi drivers without computerized service strategy.
- The average net revenue of case of 1000 electric taxis with 30 kWh battery using mode 3 charging is ranked top 0.4% among traditional taxi drivers without computerized service strategy.

The results shows that the optimal policy of MDP can enable electric taxi drivers to make comparable revenues as traditional taxi drivers.

2) Carbon Emission Evaluation:

Setting: Besides of net revenues as economic motivation, an important benefit is the reduction of carbon emission by switching from ICE taxis to electric taxis. Although electric taxis do not produce tailpipe emissions, the electricity grid to recharge the battery may still produce emissions. In this section, we estimate the CO₂ emission of electric taxis, as compared with ICE taxis, with computerized service strategy optimization. The CO₂ emission factors of electricity and gasoline are obtained from eGrid of Long Island [2]:

- Emission factor of electricity: 0.7007 kg/kWh
- Emission factor of gasoline: 2.348 kg/liter

Observations: We consider taxis working in all shifts. Fig. 14a shows the daily energy consumption of 1000 taxis for morning shifts, while Fig. 14b shows the daily energy consumption for night shifts. We use miles per gallon gasoline equivalent to convert the consumed gasoline to kWh (i.e., 1 gallon of gasoline equals to 33.7 kWh). Fig. 14c shows the annual energy consumption of different numbers of electric taxis. We observe that ICE taxis consume around 4 times more energy than electric taxis. Fig. 14d shows the corresponding CO₂ emissions of different numbers of electric taxis. We observe that up to 15 thousand metric tons CO₂ (equal to 1560 home’s energy use for one year) can be saved by replacing 1000 ICE taxis by electric taxis.

VI. CONCLUSION

In this paper, we employ Markov Decision Process to model computerized taxi service strategy and optimize the

strategy for taxi drivers considering electric taxi operational constraints. We evaluate the effectiveness of the optimal policy of Markov Decision Process using a big data study of real-world taxi trips in New York City. The optimal policy can be implemented in an intelligent recommender system for taxi drivers. This becomes more viable especially due to the advent of autonomous vehicles. Our evaluation shows that computerized service strategy optimization allows electric taxi drivers to earn comparable net revenues as ICE drivers, who also employ computerized service strategy optimization, with at least 50 kWh battery capacity. Hence, this sheds light on the viability of electric taxis.

REFERENCES

- [1] NYC Taxi & Limousine Commission. (2014) Taxicab Fact Book.
- [2] USA Environmental Protection Agency. (2017) Greenhouse Gases Equivalences Calculator.
- [3] South China Morning Post. (2017) After Hong Kong Failure, China’s BYD Joins Singapore Launch.
- [4] E. Wilhelm, J. Siegel, S. Mayer, L. Sadamori, S. Dsouza, C.-K. Chau, and S. Sarma, “Cloudthink: A scalable secure platform for mirroring transportation systems in the cloud,” *Transport*, vol. 30, no. 3, 2015.
- [5] D. Zhang, L. Sun, B. Li, C. Chen, G. Pan, S. Li, and Z. Wu, “Understanding Taxi Service Strategies From Taxi GPS Traces,” *IEEE Trans. Intell. Transp. Syst.*, vol. 16, pp. 123–135, 2015.
- [6] S. Liu, Y. Yue, and R. Krishnan, “Non-Myopic Adaptive Route Planning in Uncertain Congestion Environments,” *IEEE Trans. Knowledge and Data Engineering*, vol. 27, pp. 2438 – 2451, 2015.
- [7] S. Liu and S. Wang, “Trajectory Community Discovery and Recommendation by Multi-Source Diffusion Modeling,” *IEEE Trans. Knowledge and Data Engineering*, vol. 29, pp. 898–911, 2017.
- [8] J. Zhao, Q. Qu, F. Zhang, C. Xu, and S. Liu, “Spatio-Temporal Analysis of Passenger Travel Patterns in Massive Smart Card Data,” *IEEE Trans. Intell. Transp. Syst.*, vol. 18, pp. 3135–3146, 2017.
- [9] J. Yuan, Y. Zheng, L. Zhang, X. Xie, and G. Sun, “Where to Find My Next Passenger?” in *ACM Int. Conf. Ubiquitous Computing (UbiComp)*, 2011.
- [10] M. Qu, H. Zhu, J. Liu, G. Liu, and H. Xiong, “A Cost-effective Recommender System for Taxi Drivers,” in *ACM Int. Conf. Knowledge Discovery and Data Mining (SIGKDD)*, 2014.
- [11] H. Rong, X. Zhou, C. Yang, Z. Shafiq, and A. Liu, “The Rich and the Poor: A Markov Decision Process Approach to Optimizing Taxi Driver Revenue Efficiency,” in *ACM Int. Conf. Information and Knowledge Management (CIKM)*, 2016.
- [12] C.-M. Tseng and C.-K. Chau, “Viability Analysis of Electric Taxis Using New York City Dataset,” in *ACM Workshop on Electric Vehicle Systems, Data and Applications (EVSys)*, 2017.
- [13] —, “Personalized Prediction of Vehicle Energy Consumption based on Participatory Sensing,” *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 11, pp. 3103–3113, 2017.
- [14] C.-K. Chau, K. Elbassioni, and C.-M. Tseng, “Drive Mode Optimization and Path Planning for Plug-in Hybrid Electric Vehicles,” *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 12, pp. 3421–3432, 2017.

- [15] T. Carpenter, A. R. Curtis, and S. Keshav, "The Return On Investment for Taxi Companies Transitioning to Electric Vehicles - A Case Study in San Francisco," *Journal Transportation*, vol. 41, pp. 785–818, 2014.
- [16] New York government. (2017) Electric Vehicle Charging Stations in New York.
- [17] A. Furiere. (2017) The Gaia-SINS federated projects.
- [18] FleetCarma. (2014) Real-world range ramifications: heating and air condition.
- [19] C. Bingham, C. Walsh, and S. Carroll, "Impact of Driving Characteristics on Electric Vehicle Energy Consumption and Range," *IET Intelligent Transport Systems*, vol. 6, no. 1, pp. 29–35, 2012.
- [20] L. Feng and B. Chen, "Study the Impact of Driver's Behavior on the Energy Efficiency of Hybrid Electric Vehicles," in *ASME Intl. Design Engineering Technical Conference*, 2013.

ACKNOWLEDGMENT

The authors would like to thank Srinivasan Keshav and Sgouris Sgouridis for helpful suggestions and discussions.