Off-Policy Evaluation of Driver Repositioning Policies

Columbia Data Science Capstone Project - Spring 2021

Danyang Han (dh2990@columbia.edu), Anita Pinto (ap3650@columbia.edu), Elizabeth Yum (eyy2106@columbia.edu)
Mentors: Vineet Goyal (vgoyal@ieor.columbia.edu), Zhiwei Qin (qinzhiwei@didiglobal.com)

***Abstract*—This paper describes an approach to tackle the problem of driver repositioning on ride-hailing platforms through off-policy evaluation by the creation of an environment simulator that leverages historical data to stochastically generate demand, passenger/driver matching and destination probability to be used in off-policy evaluation, with a particular focus on the uncensoring of ridership demand.**

***Index Terms*—off-policy evaluation, driver repositioning, reinforcement learning, uncensored demand, taxi optimization**

I. Introduction

In large, highly competitive taxi markets, taxi drivers' incomes are highly dependent on their ability to be able to find their next passenger quickly and can vary widely depending on their working behaviors and preferences – when and where to start a shift and their repositioning behavior between rides. The time spent cruising for new passengers present greater problems than just that of opportunity cost in lost income for taxi drivers. Empty, cruising vehicles contribute to greater congestion and wasted resources, and in a recent study commissioned jointly by the New York City Taxi and Limousine Commission and Department of Transportation [1], the cruising rate of drivers in Manhattan core (below 96th street) was 41%. In addition, many studies have been conducted on pattern recognition and reinforcement learning to minimize driver idle time and maximize driver income, but most efforts have been limited to evaluation of historical data.

In this paper, we aim to tackle the problem of driver repositioning on ride-hailing platforms through off-policy evaluation by the creation of an environment simulator that leverages historical data to stochastically generate demand, passenger/driver matching and destination probability to be used in off-policy evaluation. Of particular note is the issue of how to effectively back into total demand given the presence of only censored demand data, as historical rides only contain passenger rides that were fulfilled.

II. Related Work

To our knowledge, no equivalent research has been published that aims to create an off-policy evaluation process through the creation of an environment simulator.

The most critical piece of the environment simulator is an effective way of generating demand solely on historical observations which are effectively censored as they cannot reflect unfulfilled demand. Research on censored demand that tackles the newsvendor problem may have some parallels. In the newsvendor problem, a newsvendor seeks to optimize stock to maximize profit under fluctuating demand. Research from Mersereau [2] approached the problem of estimating sales when sales are a censored representation of the underlying demand process and the impact of inventory record inaccuracy in an environment with censored observations. Within this paper, we describe our method of uncensoring demand following Mersereau's approach of gradient descent for a discrete distribution such as the Poisson distribution.

Research on optimal routing of a vacant taxi includes work from Yu et al. [3] proposing a process of expressing the routing problem as a Markov Decision Process. States were defined by the node at which a vacant taxi is located and actions were taken as the link to take out of the node. State transition probabilities were expressed as passenger matching probabilities and passenger destination probabilities, and the MDP problem was solved through value iteration. Within this paper, we describe our approach toward adapting Yu et al.'s approach toward the problem of passenger matching probabilities.

III. Methodology

Figure 1 below depicts at a high level the proposed overall process.

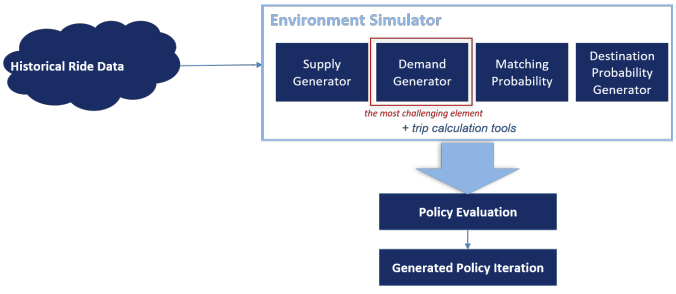

Fig. 1. Our Methodology

First, historical data is leveraged to help create an environment simulator that generates supply, demand, matching probability and destination probability by taxi zone and time.

The most challenging piece of the environment simulator is the demand generator. We are restricted to use of NYC Taxi & Limousine Company (NYC TLC) publicly available data which is confined solely to observed, captured demand. However, successful implementation of a full environment simulator hinges on the ability to estimate total demand including both captured and uncaptured, and therefore unobserved data.

The outcome of the environment simulator is combined with trip calculation tools that generate the fare, travel time, and trip distance for each simulated taxi trip. The simulations are then fed through a Policy Evaluation process that updates values at each state based on the outcome of the Monte Carlo simulation. Finally, it is possible for the results of policy evaluation to be fed through a generalized policy iteration process.

IV. Dataset

A. Description

For historical ride data, we are leveraging data from the NYC TLC, the agency responsible for regulating New York City's medallion (yellow) taxis, street hail livery (green) taxis, for-hire vehicles (FHVs), commuter vans, and paratransit vehicles. The TLC receives taxi trip data from the technology service providers (TSPs) that provide electronic metering in each cab for each medallion taxi. It then collects this and data from FHVs and publishes data for all taxi and FHV trips completed by licensed drivers and vehicles on a monthly basis.

For the purposes of this study, we are solely leveraging Yellow Taxi trip data from June 2013, due primarily to barriers to using broader or more recent data. We are constraining the data solely to Yellow Taxi data as Yellow taxis are the only vehicles permitted to respond to a street hail from a passenger in all five boroughs, and were the dominant mode of taxi and limousine transportation in 2013. As such, the observable behavior in Yellow Taxi data is unconstrained compared to that of Green Taxis (which may only respond to street hails in select locations), and captures nearly all of the passenger ride services trips in 2013. We are limiting ourselves to data from 2013 as historical Yellow Taxi data included individual Taxi identifiers which enables identification of individual driver behavior, movements, and outcomes, which more recent Yellow Taxi data lacks. Additionally, the rise and increasing popularity of ride hailing apps means that a significant proportion of ridership in recent years is through FHV services which are subject to lower data reporting standards and missing critical data fields compared to Yellow Taxi and Green Taxi services.

B. Data Cleaning

Similar to work done in the prior iteration of the capstone, the following data cleaning steps were applied to remove abnormal data:

- Trip time: the duration of a single trip. We dropped trips with traveling time that is less than 1 minute or longer than 3 hours as most of these records have extremely short or long travelling time.
- Trip distance: the traveling distance of a single trip. We calculated the average speed of each trip based on trip distance and trip time, and drop records that have speed over 50 mph, which is the highest speed limit in New York. Trips that are longer than 30 miles are also removed because they are about twice as long as the length of Manhattan
- Trip fare: the total money a driver earns from a single trip. We dropped records with trip fare that is higher than $150, which is the trip fare of travelling 60 miles.
- Pick-up and drop-off locations: We dropped the trips whose pick-up or drop-off location were not in NYC.
- Trip pick-up and drop-off times: We removed trips with a null pick-up or drop-off time which had been defaulted to a value of 0

The total amount of data removed from the data cleaning steps applied above was 3% of all trip data (511,647 of 14,385,456 trips were removed leaving 13,873,809 total trips available for analysis), and we do not believe the data cleaning steps above introduces any undue bias.

C. Discretization and Transformation

Continuous spatiotemporal data can present particular challenges for reinforcement learning which were resolved in our project through the discretization of continuous state spaces for manageability.

Pick-up and drop-off locations were discretized by applying a transformation to map pick-up and drop-off GPS coordinates to one of the TLC's 263 Taxi Zones following the approach from the previous capstone [4]. A geographic depiction of all NYC TLC Taxi Zones can be found in Figure 2 below.

Taxi zones as defined the TLC are roughly based on NYC Department of City Planning's Neighborhood Tabulation Areas (NTAs) and are meant to approximate neighborhoods. As can be seen in the figure above, taxi zones are not of uniform size and characteristics. Rather, taxi zones in more populous areas tend to be smaller in area while taxi zones in less densely populated areas tend to be larger. The non-uniformity of taxi zones did not present many problems in our data analysis, although it did present a mild challenge in imputing traveling distances and times between zones for which data was not available. This is described further in Section V.E. Trip Calculation Tools.

For certain calculations such as for the simulation of demand, we also discretized time into 30 minute intervals. We assumed 24 hour days were split by a time interval of 30 minutes and therefore each day comprised of 48 30-minute time intervals. For example, in the instance of calculating demand for each taxi-zone, a λ parameter for the Poisson distribution was estimated for each time interval and taxi zone.

In addition to the discretizing steps described above, data was also inspected visually across a number of dimensions for understanding how best to represent state spaces and run our simulation in a way that effectively models the real world.

One observation was that there are significant differences in passenger and driver behavior between weekdays and

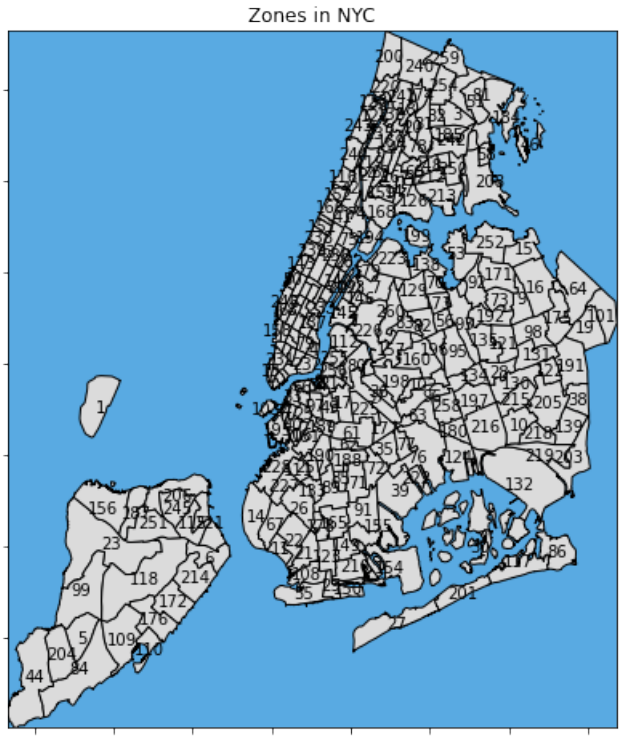

Fig. 2. TLC Taxi Zones

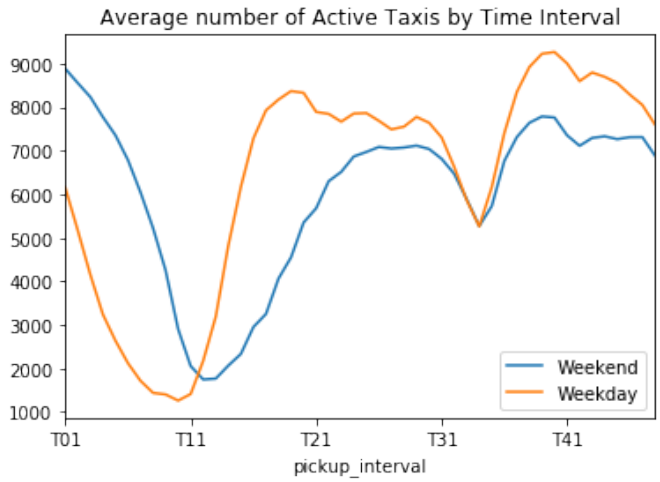

Fig. 3. Number of Taxis by Time Interval where the y-axis represents an average day broken up by 48 time intervals

weekends. For example, during weekdays, total ridership and available taxis were higher. Relative demand for taxi rides to taxi zones associated with traditional business districts were higher on weekdays compared to weekends when relative demand for taxi rides to taxi zones associated with leisure. Therefore a weekday/weekend indicator was added to the state space.

Another observation was that the majority of NYC TLC drivers work either one of two shifts – a morning shift or an evening shift of approximately 11 hours.

The figure above shows how this pattern results in two distinct peaks and troughs of number of active taxis throughout the day. Our simulation runs as a taxi driver beginning and completing a single 11-hour shift to better mimic real world behavior.

V. Components of the Environment Simulator

A. Supply Generator

For generating supply, there were two approaches that were considered. One was that of having a single agent learning within a broader environment that is simulated by historical data. The second approach was that of having multiple agents learning and following a specified policy within a dynamically changing environment. A comparison of the two policies considered is below.

Due to the computational complexity of a multi-party model, for the initial productionizing environment, it was decided to go with a single learning agent. Within the single agent framework, the supply would be determined as follows.

TABLE I
Supply Generation Options Considered

	Option 1: Single Taxi	**Option 2: Multiple Taxis**
Description	A single agent learning within a broader environment that is simulated by historical data	Multiple agents learning and following a specified policy
Pros	Computationally simple table copy	The ability to dynamically change the environment as multiple agents learn at once
Cons	Static environment, assumption that only a single agent is learning	Computational complexity

The single agent will begin either a morning or evening shift on the weekday or weekend consisting of 11 hours. The single agent would be assigned a starting location randomly. From there, the single agent would move according to the steps outlined in Section VI.B Simulator and the Environment below.

B. Demand Generator

As the most critical piece of the environment simulator, Demand generation had a number of challenges we attempted to work through.

Challenge #1: Demand is highly sensitive to zone, time, and day:

As illustrated in Figure 4, demand varies highly by zones and by day of the week. Additionally, while not explicitly depicted below, demand is also highly sensitive to the time with demand peaking for most zones within the typical "rush hours" (early AM and early evening coinciding with typical work schedules). While not explicitly a methodological challenge, any approach to demand estimation will need to be

appropriately sensitive to the differences in demand across zones, times and days.

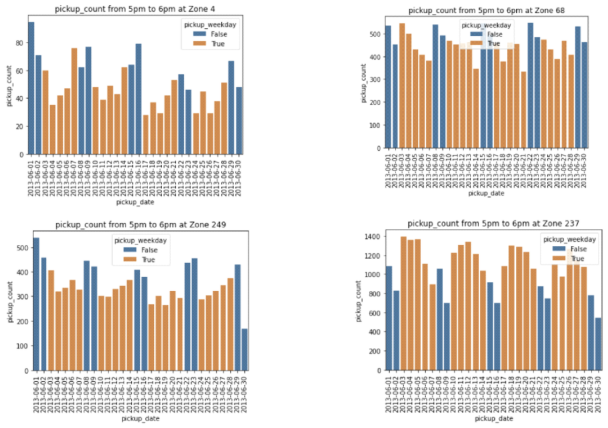

Fig. 4. Pickup counts 5-6 PM sampled across 4 Zones

Challenge #2: Historic data only contains observed demand

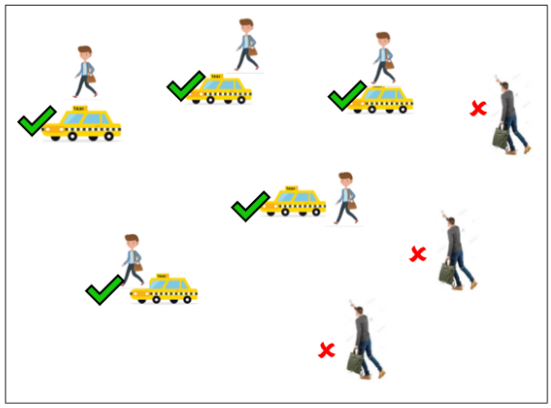

Fig. 5. Uncaptured Demand

In contrast to the newsvendor problem, directly observable data from NYC Yellow Taxi data lacks the ability to directly observe when demand exceeds available supply. In the newsvendor problem, sellers have history about sales (s_t) and stock (j_t), and also record a stockout indicator $c_t = 1$ when $s_t = j_t$, and $c_t = 0$ when $s_t < j_t$. In contrast, for Yellow Taxi data, which are continuously in supply, but with supply distributed unevenly across New York City (an empty taxi does not benefit a potential rider if it is 10 or even 1 block away), it is difficult to approximate when demand has been lost.

The demand is the passenger arrival rate arriving at a particular taxi zone in the time interval. The passenger arrival rate is assumed to follow a one-dimensional space-time Poisson process with rate λ. In general practice, the demand rate λ is often approximated by observed demand rate. The problem of estimating true demand given censored lost demand has been studied extensively [2], however as mentioned above, the challenge with the NYC trip data was to estimate the censoring indicator.

1) Estimating Censoring Indicator: We considered the following parameters for estimating the censoring indicator.

- **Combination of inter-pickup times and frequency:** For calculating inter-pick up times and the frequency of drop-offs within each taxi zone, we computed the inter-pickup time as the difference between the prior trip's drop off time and the next trip's pick up time and the number of drop offs within each 15 minute interval.
- **Price Surge:** We explored existing data to look at fare patterns over 15 minute intervals in the figure below. However, as Yellow Taxis go by rates which do not have a directly observable price surge indicators for periods of high demand as many FHV services would have, variation was limited.

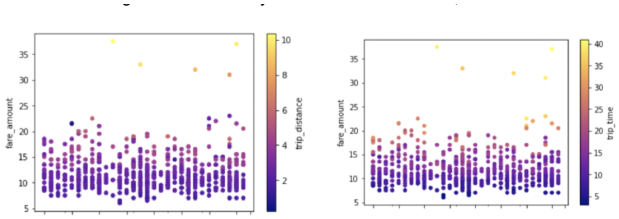

Fig. 6. Price Surge

- **Net taxis in zone:** Calculated as number of (taxis that dropped off in t-1) - number of (taxis that picked up in t-1). This option was discussed, but it would need to be customized by state and by zone, and it was unclear what the cut-off point should be.
- **Taxi utilization:** The percentage of available taxis in t-1 that picked up in time t.Again it was unclear what the cut-off should be, low utilization places are possibly the sources of highest lost demand (example, Queens).
- **Inter-Pick Time:** This is the time difference between two pick up intervals within a zone. The longer intervals indicate longer wait for taxis and possible lost demand. However we noticed that for a high traffic busy zone, the inter-pickup arrival time was near to null.

We used a combination of inter-pickup intervals and the Taxi utilization information to arrive at estimating censoring indicators for each of the taxi zones and time intervals. To keep the censoring flag independent of taxi zone, time interval, and day of the week, this parameter was estimated for each of the taxi zones and time intervals independently.

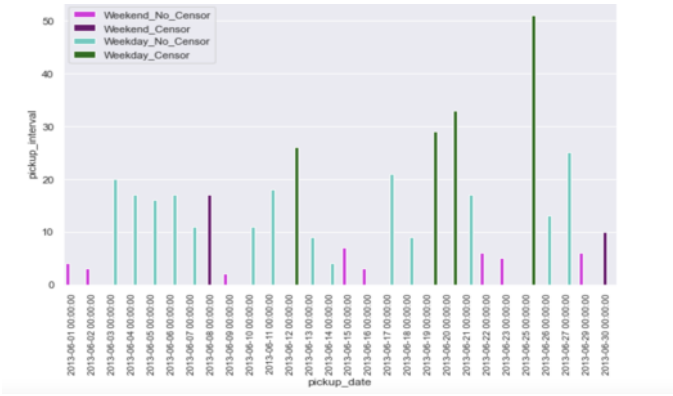

Fig. 7. Pick up Intervals Zone 36 (Brooklyn Bushwick North) at 12 Midnight

Figure 7 above shows how the censoring indicator calculated for a taxi zone 36 (Brooklyn Bushwick North) for time interval 12:00-12:30 based on inter pick up interval times. Since the censoring flag is independent for weekday/weekend, notice how an inter-pick interval for weekday would be considered no censoring however the same range of values for weekend would indicate censoring.

2) Calculation of Uncensored Demand: Generally, the observed arrival rate for a taxi zone at a given time interval, λ would be the mean of the 30 days arrival observations, x_1, x_2, .. , x_{30}

However, for estimating the true demand inclusive of the censored demand, we set the demand estimate λ as the maximum likelihood parameter estimate given the censoring indicators for each of these arrival observations (x_1,δ_1), (x_2,δ_2), ... (x_{30},δ_{30}), where,

$\delta_i = 0$ indicates no censoring

$\delta_i = 1$ indicates censoring

Figure 8 below shows the censor indicators for different taxi zones at Rush Hour vs Off Peak hour. Notice how the censoring indicators vary across zones and time intervals.

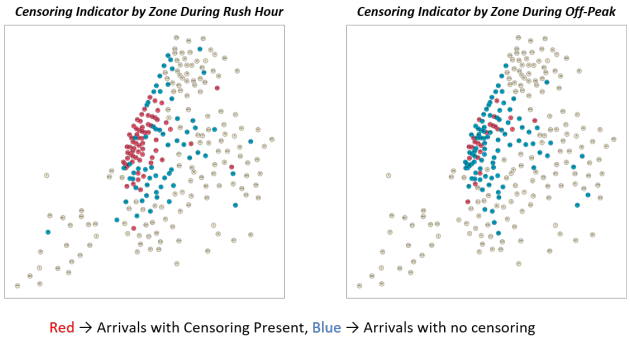

Fig. 8. Comparison of Censoring by Time

The total uncensored demand parameter λ was be estimated using:

$$\lambda_{ML} = argmax(\sum_{i=1}^{n} (1-\delta_i)\ log(f(x_i,\lambda) + \delta_i log(1-F(x_i,\lambda)\ )$$

Where,

$$f(x_i,\lambda) = \frac{e^{-\lambda}.\lambda^{-x_i}}{x_i!}.$$

$$F(x_i,\lambda) = \sum_{i=1}^{n} \frac{e^{-\lambda}.\lambda^{-x_i}}{x_i!}$$

C. Matching Probability

To accurately model the passenger search behavior of vacant taxi drivers, we developed a framework to calculate matching probability similar to the approach in Yu et al [3], which was based on taxi movement on a given road network. The matching probability is a combined probability of each of these individual probabilities for a taxi zone given a time interval Δt.

1) Probability of at least one pick up at taxi zone in Δt time interval : The arrival rate is assumed to follow a Poisson distribution with parameter λ which is the newly estimated uncensored demand rate. The probability of at least one pick up in a given interval, would be 1 - Probability of no arrivals for that time interval Δt.

$$\text{Pr[At least one Arrival]} = 1 - e^{-\lambda}$$

Where,

λ is the estimated total uncensored demand parameter

2) Probability of no vacant taxi at a taxi zone in the given time interval: To model this probability, we used the approach in Yu et al [3], where the probability of no vacant taxi could be estimated using the drop-off counts occurring in that taxi zone for a given interval. We assume the drop offs also follow a Poisson distribution with parameter λ_d which could be estimated as the average of all drop-offs occuring in that taxi zone in that time interval Δt.

$$Pr[X_d = 0] = e^{-\lambda_d}$$

Contrary to the Yu et al [3] approach, the probability of vacant taxi calculated in the NYC taxi data couldn't provide an accurate representation of the probability of no vacant zones in a zone. The area of consideration in our approach is a wider one compared to the Yu et al [3] approach where the granularity of probability calculations was confined to nodes (for which they estimated a total of 13,531 nodes in the Shanghai roadway network separated into 4,518 zones) instead of the whole taxi zone. The challenge we encountered was, for very busy zones the drop off counts in a given time interval Δt of 30 min was always high and hence the probability of no vacant taxi would always drop the probability to a low value. A better approach would be to reduce the area of consideration and the time interval used to estimate the drop-offs.

To obtain a modified Probability of no vacant taxi in a zone, we made use of the difference in drop-off counts and pick-up counts occurring at a taxi zone in a given time interval Δt.

Pr[NoVacantTaxi]= $1 - ((C_d - C_p)/C_d)$ if $C_d > C_p$ else 1

Where,

C_d = Number of Drop-offs.

C_p = Number of Pick ups

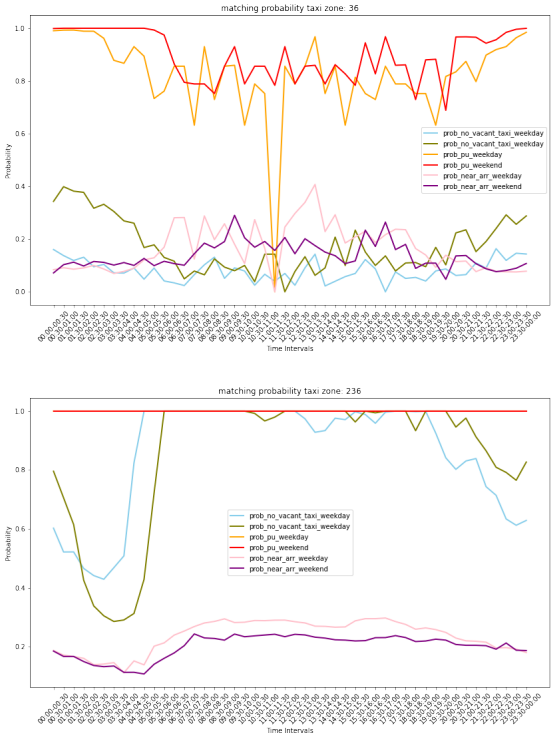

Fig. 9. Matching Probability for Taxi Zone 36(Top) and Taxi Zone 170(Bottom)

3) Probability of zone being the nearest one to an arrival: The probability of being the nearest zone to an arrival would simply be the ratio of arrivals in that zone to the total arrivals in that zone and its adjacent zones. This probability gives an estimate of the incentive to reposition to an adjacent zone given the arrival rates of all the adjacent zones.

$$\text{Pr[nearest to arrival of passenger]} = \frac{\lambda_z}{\sum \lambda_i}$$

where,

λ_z is the arrival rate at the zone

$\sum \lambda_i$ is sum of arrival at all adjacent zones and the zone

Similar to the probability of no vacant taxi discussed above, the calculation of this probability would be more appropriate at a node level than a whole taxi zone level. We noticed that for a very busy high traffic zone, this probability brought down the matching probability depending on the arrival rates of adjacent zones. Ideally, this probability gives a better approximation when used for smaller areas under consideration instead of the entire taxi zone. Nevertheless, we used this probability calculation for the reposition algorithm in the environment simulator. To add some randomness to the reposition algorithm and differentiate zones with nearly equal arrival rates, we used softmax function on the probability calculation of nearest arrival.

$$\text{Pr [near arrival from } z_a \text{ to adjacent zones } z_i] = \frac{e^{(z_i - z_a)\cdot\beta}}{\sum_{adj} e^{(z_i - z_a)\cdot\beta}}$$

Altogether, the figure on the opposite page shows the all individual components of the probabiltiy matching calculations for Zone 36, Brooklyn Bushwick North, and zone 236, Manhattan Upper East Side North.

The plots in Figure 10 and Figure 11 show the heat map distribution of the probability matching values across NYC taxi zones. Notice how the Manhattan regions have higher probability matching than other taxi zones. It is also understandable that since this is June 2013 data, the proportion of rides with Uber/Lyft might be significantly lower compared to the current proportions, so a heatmap distribution of more a recent taxi data might show some variance than what is observed. It is also interesting to observe that the heatmap distribution for shift B, i.e., evening shift shows a wider spread than the morning shifts for both weekday and weekend data.

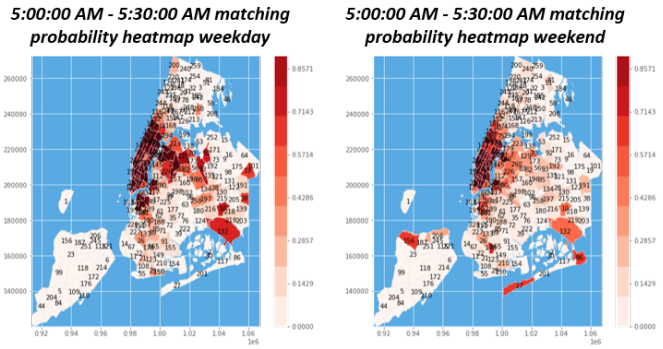

Fig. 10. Matching Probability Heatmap for NYC Taxi Zones at 5AM

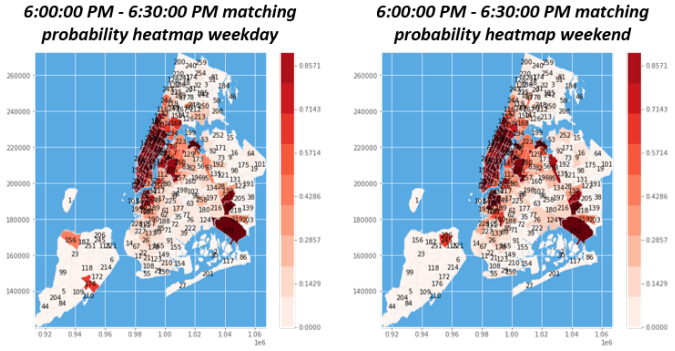

Fig. 11. Matching Probability Heatmap for NYC Taxi Zones at 6PM

D. Destination Probability

Destination probability was determined independently from other components such as demand. To determine destination probability, we reviewed historical data by shift and weekend/weekday to create a Markov matrix of historical rider demand based on their point of origin. This Markov matrix was then applied after driver matching based on the driver's pick-up location to determine the passenger's destination. As an example, the figure below depicts sensitivity to destination at the shift-level on weekdays. The first heatmap shows top destinations by total number of rides during the morning shift where demand is heavily concentrated in the Midtown area. The second heatmap shows top destinations during the evening shift when demand is much more evenly dispersed and with more activity in zones associated with traditionally residential and leisure neighborhoods.

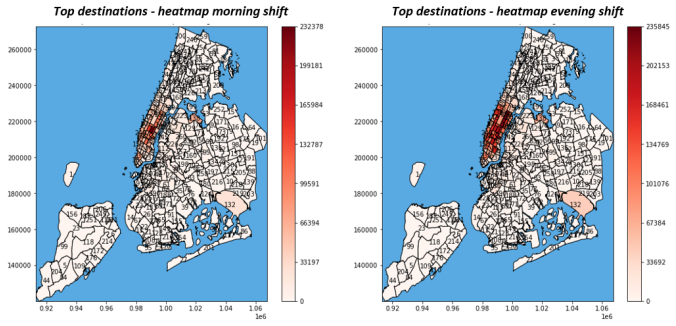

Fig. 12. Heatmap of Top Destinations by Shift on Weekdays

However, there are instances in which demand was determined to be censored, and historical destinations observed may not accurately reflect total destinations demanded had uncensored demand been taken into account. The destination probability could be extended further to incorporate more randomness. For example, in instances when demand is censored, we could apply a 90% chance the passenger will demand a destination that has been previously observed from that point of origin, and a 10% chance the passenger will demand a new destination selected at random or selected from an expanded list of destinations from nearby neighbors. This extension, however, would require additional work to determine the best approach, and would require the calculation of trip durations and fares between location pairs that have not been observed in historical data. While aerial distance between zones is easily available information we have on hand, this does not always easily translate to travel time or trip fares. For example, trip fares are a function of not just distance, but also average speed between the two locations, which can be sensitive to times of high traffic, and tolls encountered (or avoided at the expense of distance and time) and therefore, fares cannot always be easily imputed from distance alone. The estimation of accurate fares between any two points would require a small model with additional inputs beyond aerial distance or the assumption of an address/coordinates and usage of a map/directions API.

E. Trip Calculation Tools

In addition to each of the major components of the environment simulator described, we also used historical data to calculate underlying trip metrics to feed into the simulator, and used reasonable assumptions where historical metrics were not available.

Historical data was used to calculate the following:

- Average trip duration between all pick-up and drop-off location pairs observed
- Average trip mileage between all pick-up and drop-off location pairs observed
- Average fare between all pick-up and drop-off location pairs observed

In addition to historical data, we also supplemented the simulator with a few assumed and imputed values:

- Following the approach from the prior semester [4], we estimated fuel cost as distance·fuel unit price. The fuel unit price was calculated by using the average gas price in 2013 of $3.602 per gallon and the taxi fleet average fuel economy of 29 MPG (miles per gallon) [5]. Thus, the estimated fuel unit price in 2013 was $0.124 per mile
- When drivers cruised for a new passenger after dropping off a passenger, we assumed an average cruising speed of 10 mph. For reference, the average weekday travel speed in Midtown Manhattan was estimated as 6.1 mph in November 2010 [1]. The top speed limit on New York City highways is 50 mph or can be 65 mph on certain routes between Newark Airport and New York City, but when drivers are cruising for passengers they are typically on local roads. The speed limit on local roads can vary, but if not posted, the default speed limit is 25 mph[1] which combined with stop lights, stop signs, and modified driver behavior as they search for a customer rather than drive with a destination in mind should bring the average cruising speed down close to our assumed speed.
- Depending on the repositioning strategy assumed, drivers may travel to a neighboring zone in search of passengers. Where ride data between neighborhoods was not observed, the coordinate data of each zone was used to calculate the distance between midpoints of each zone and this aerial distance was multiplied by a factor of 1.2 to take into account that actual roads and therefore travel distance are rarely strictly linear. In certain cases, the model could output an order for the taxi driver to reposition in their current zone. In that case the length of the taxi zone, representative of the linear length of all its boundaries, was taken and converted to miles and used to impute an approximate measure of half of the length of the taxi zone as represented as a polygon.
- Additionally, when repositioning to a zone to which ride data is not available, we also imputed the trip duration based on the distance above (the aerial distance·1.2) and assuming a cruising speed of 10 mph. Unlike the

[1] https://portal.311.nyc.gov/article/?kanumber=KA-01368

problems described above with calculating trip durations and fares between new location pairs, in the case of a driver repositioning to a new zone, the repositioning is constrained to the nearest neighboring zones. As there are fewer sources of variation for calculating average duration between neighboring zones and there was no need to calculate trip fare, we felt comfortable with making assumptions based solely on aerial distance.

VI. Policy Evaluation Process

One of the advantages of our simulation approach is the ability to evaluate any given reposition policy, and therefore go "off-policy" relative to historical data. In this section, we demonstrate policy evaluation on three different reposition policies.

A. The Policies Evaluated

1) Hot Spot Tracing (Deterministic): Reposition to a zone with the highest arrival rate among a group of adjacent zones of the current zone. For example, at time t, if a driver's current zone has adjacent zones $1, \ldots,$n. Then compute the historical demand of the zone itself and its adjacent zones at time t. Denote the historical demand of current zone as δ_0, and that of the adjacent zones to be δ_1, ... , δ_n. Find $max(\delta_i)$ and reposition taxi to zone i.

2) Hot Spot Tracing (Probabilistic): Uses softmax to normalize demands of adjacent zones into a probability distribution. Extending the example above, for i in range $0...n$, we computed

$$P(i) = \frac{exp[\beta * (\delta_i - \delta_0) - \epsilon]}{\sum_{i=0}^{n} exp[\beta * (\delta_i - \delta_0) - \epsilon]}$$

Note that $\sum_{i=0}^{n} P(i) = 1$. We then repositioned using this probability distribution

3) Random Reposition Policy: Reposition randomly to an adjacent zone.

B. Simulator and the Environment

The following is a step-by-step description of how the simulator is initialized and how it runs.

- Setup
  - ◇ Driver object maintains driver's spatial and temporal information, as well as the earnings of a shift
  - ◇ Environment takes parameter weekday or weekend. It maintains information calculated from historical data.
    - ∗ Given with zone and time, environment returns matching probability, and transition or destination probability.
    - ∗ Given with a start zone and a destination zone, environment returns travel distance, trip duration, cost of fuel and revenue from fare.
    - ∗ Given with a zone, environment returns a list of adjacent zones

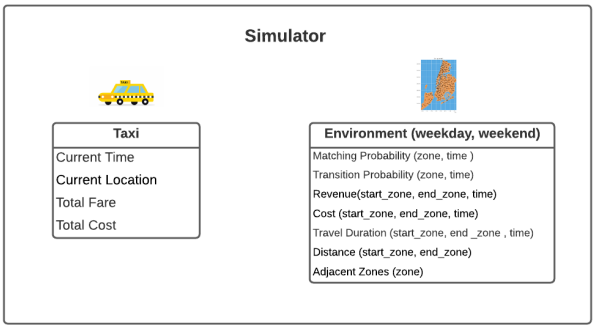

Fig. 13. Simulator Block Diagram

- Initialization
  - ◇ Driver is initialized at
    - ∗ time = beginning of a shift (a datetime object (yyyy/mm/dd/hour/minute/second))
    - ∗ starting zone = any of the 263 zones
    - ∗ total fare = 0
    - ∗ total cost = 0
  - ◇ Initialize environment to be either a weekday environment or a weekend environment
- Simulation of one episode
  Figure 14 illustrates the process of simulating one shift.
  Repeat the following steps, until the end of the shift:
  Stage1: decide whether a taxi will get a pick up at current location.
  Stage2-1: If so, find the destination zone and update driver attributes with new trip information calculated from environment. Go back to stage 1.
  Stage2-2: If not, reposition taxi driver using a specific policy. Update driver attributes with reposition trip information from environment. Go back to stage 1.

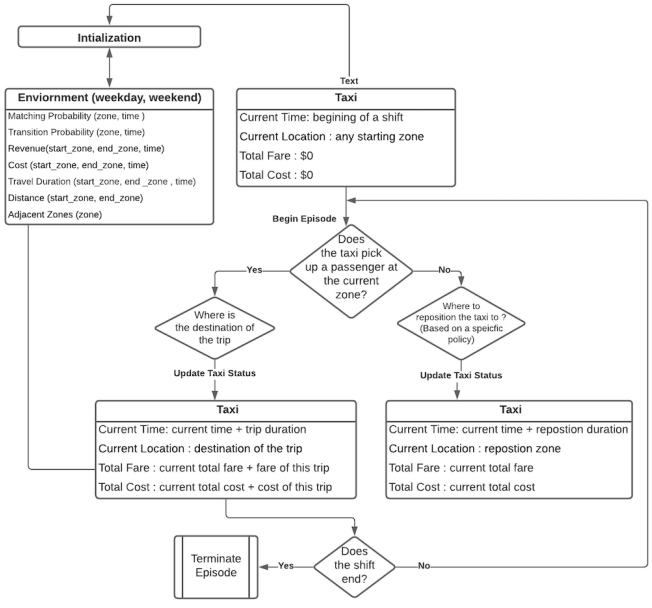

Fig. 14. Environment Simulator Data Flow

- Driver Trajectory
  Figures 15 and 16 show driver trajectories starting from different taxi zones for weekday and weekend and shift A (morning and afternoon) and B (evening and early morning) variants. The red dashed lines, indicate driver repositioning which is controlled and guided by the policy under evaluation. The blue lines indicate actual passenger trips taken by the driver(taxi) during the simulation. The destination taxi zone is controlled by the Destination Probability matrix derived from the historical trip data. The matching probability decides if a driver gets a ride or not after reaching a new taxi zone. As observed in the figures below, it was noticed that though our taxi zones originated from different taxi zones and different days and shift times, the simulator always managed to get the driver to Manhattan region. We come to understand that, this is mainly due to the fact that both repositioning logic and trip destinations are derived from historical data where the concentration of trips is higher in Manhattan region.

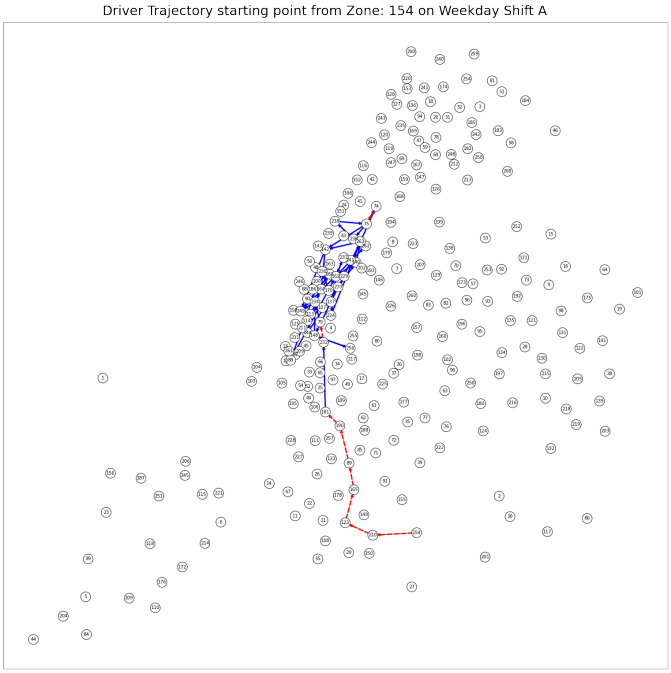

Fig. 15. Driver Trajectory Starting at Taxi Zone(154) on Weekday at 5 AM

C. Policy Evaluation

Having run multiple iterations of simulations on each of the 3 different policies, the goal is to evaluate and compare these policies. We considered Monte Carlo methods for learning the state-value function for a given policy.

Key Assumptions: As the dynamic of real street network is continuous, complex and changeable, we have made the following assumptions to make policy evaluation more feasible:

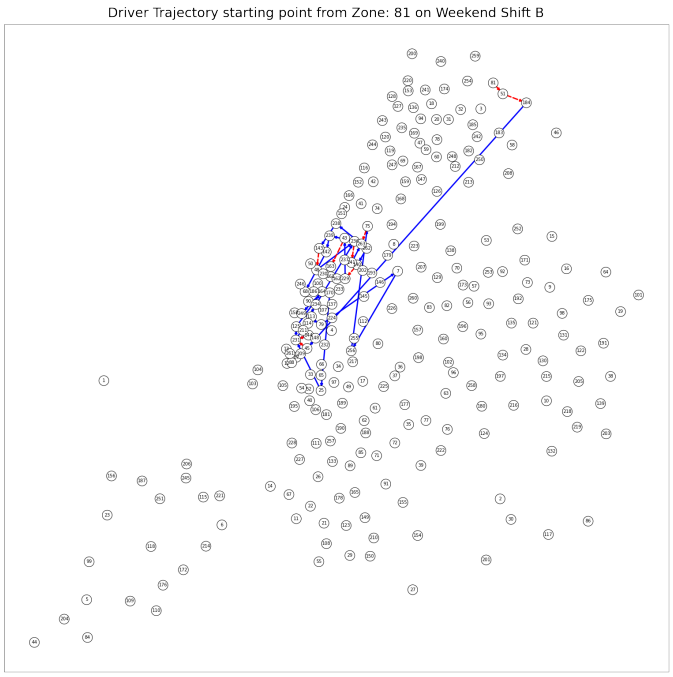

Fig. 16. Driver Trajectory Starting at Taxi Zone(81) on Weekend at 6 PM

1) Coordinates are partitioned into discrete taxi zones: The probability of receiving an order (matching probability) and probability of moving to a destination taxi zone, distance towards other taxi zones (and related fuel cost), and time spent to travel to another taxi zones are invariant within a taxi zone. In other words, taxi zone is the smallest unit recording the driver's location.

2) Time is partitioned into discrete time intervals of 30 minutes each denoted by Δt: The time space is denoted by T = [1; 2; ... ; 48].

3) Single Agent Simulation System: There is only one driver following the policy the model derives, i.e. one agent.

4) State Space S [z,t]: The state is represented by [z,t] taxi zone z and followed by time interval at which the trip started. We have in total 263(zones) x 48(time intervals) state space, and state values were computed separately by weekend/weekday and by shift.

5) Actions: There are only two actions an agent takes, one when there is successful match for a trip, the driver moves to the destination taxi zone, second one being, when the agent fails to acquire a trip, the policy decides the reposition action, which can be [a] whether the taxi cruises in the current zone or [b] moves to any adjacent zone. In this current setup, all reposition policies consider only current zones or adjacent zones.

6) Shift Time: Each shift time is 11 hours from the start of the simulation. We are not considering cases where the driver decides to terminate the shift earlier than 11 hours. There are 2 shift types, we have typically used for our evaluation, Shift A [5AM-4PM] and shift B [6PM-5AM].

Given the current setup, we chose to evaluate and compare different policies using Monte Carlo evaluation to learn and compare state value functions under different policies. The value of the state is the expected return starting from that state. Note that, state means the taxi zone and the associated time interval the trip starts. We compared the state values for weekend/weekday and shift-A/shift-B separately.

In particular, suppose we wish to estimate, the value V_s^π of a state s under policy π, given a set of episodes obtained by following and passing through s. Each occurrence of state s in an episode is called a visit to s. The MC method estimates V_s^π of state as the average of the returns following all the visits to s in a set of episodes. Note that, since t is also linked to a state with zone z, the state s[z,t] would occur only once in every episode as time progresses forward in every episode. The obvious way to estimate the state value would be to simply average the returns observed after visits to that state through multiple iterations. A depiction of how this works and the algorithm we used follows below.

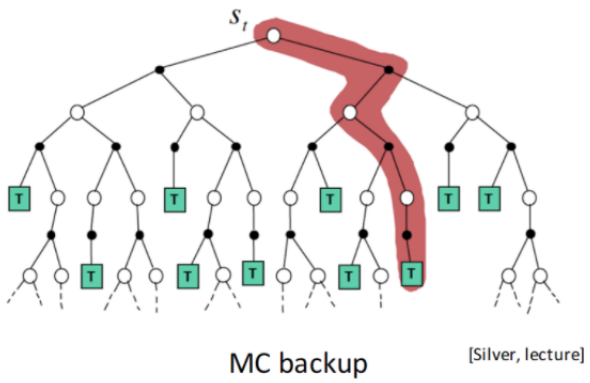

Fig. 17. State transitions in Episode

Algorithm 1: MC Evaluation

Initialize:
V[s] = 0 , V[z,t] = state value functions
Returns[s] $\rightarrow$ empty list for all returns
N[s] = 0 $\rightarrow$ state hits
Repeat forever
  Generate an episode using policy π
  Calculate G(s) = R(t+1)+R(t+2) + … R(T)
  Append G(s) to Returns[s]
  N[s] = N[s] + 1
  V[s] = V[s] + (1/N[s] * (G[s] - V[s])
  (V[s] updated using Incremental Mean)

D. Simulation Results

The simulations were initially run at random. We observed that starting in certain taxi zones like those of Staten Island showed much less earnings throughout and account for starting location as a contributor to variation in total earnings, we also brute force starting from each of the 263 taxi zones. To compare the 3 policies amongst each other, we ran the simulations at a fixed starting time for each of the shift variants, namely, 5AM for the shift A and 6PM for shift B. In total, we ran a minimum of 120K simulations (3 policies·2 indicator values for weekday/weekend·2 shifts·263 taxi zones·40 times each)

1) Observations:

- The figures below show a comparison our simulator trip earnings with those that of the historical trip earnings.

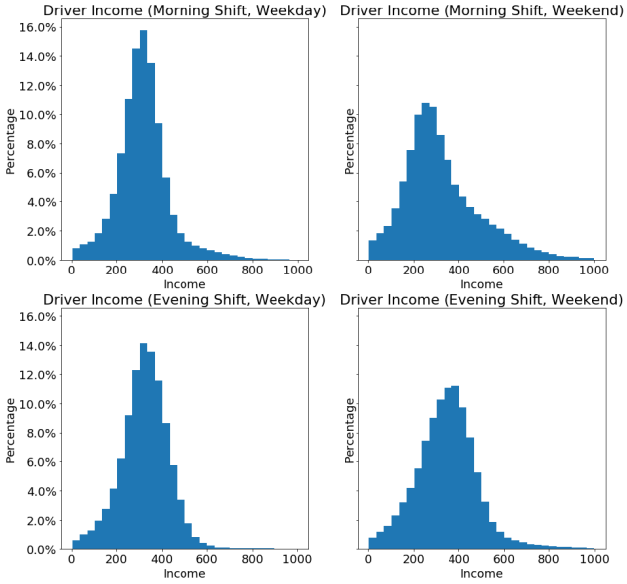

Fig. 18. Historical Trip Data (Gross) Income

Fig. 19. Simulator trip earnings

Our simulator trip earnings nearly matched the histogram earnings with average earnings around $300 to $400 per shift. However, there are two main differences of note. Firstly, the modes of our trip earnings were slightly lower than that of the histogram of gross driver income. Secondly, while our simulator seemed to mimic lower tails of driver income effective, but did not produce many total driver shifts with high rates of earnings (e.g., $600+). We are not overly concerned with these differences due to the following reasons.

Firstly, the histograms of driver income were based on gross total trip fares during their shift and are not inclusive of fuel costs for both passenger trips and cruising which were added to our simulator. With the inclusion of these in driver income (which would require many assumptions on cruising distances and speed), we believe the mode of actual driver net income would be closer to our simulator.

Secondly, we believe our simulator did not generate shift values that go well above the mean value of actuals as [a] our simulator constrained drivers to 11-hour shifts and did not allow drivers to work in excess of this for greater profits, and [b] when drivers were matched with trips, the fare that they recorded was the average fare recorded between those two destinations in historical data which precludes the ability of a driver in our simulation to have an unusually "good" day of tips.

- Comparing the earnings across the three policies, we observe that the Hot Spot Tracing repositioning policy (repositioning to the neighboring zone with the highest arrival rate) had higher earnings than the policy running on a random walk to any neighboring taxi zone.

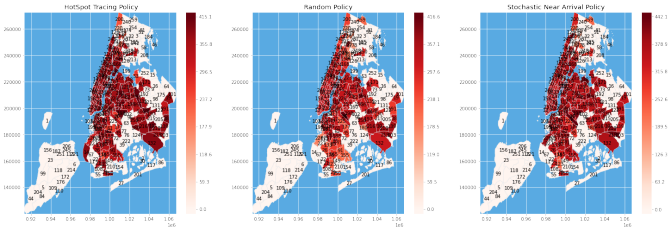

Fig. 20. policy earnings comparisons

- It was also observed that there were lot of taxi zones especially the Staten Island regions where the earnings were near to zero. This is no surprise to us because our simulator is run on passenger matching probabilities[Fig.10-11], destination probabilities[Fig.12] and nearest arrival rates all of which are comparatively low in these regions.
- To reaffirm our understanding and recheck why the Staten Island Taxi zones had near to zero earnings, we ran standalone simulations to check how our simulator behaves when a trip is initiated in one of these taxi zones. The following Driver Trajectory path confirms our understanding that because of low matching probability, destination probability and arrival rates, any trip initiated at these taxi zones, loops around repositioning itself within its neighboring zones.(Red dashed lines indicate repositioning and blue lines indicate actual passenger trip).

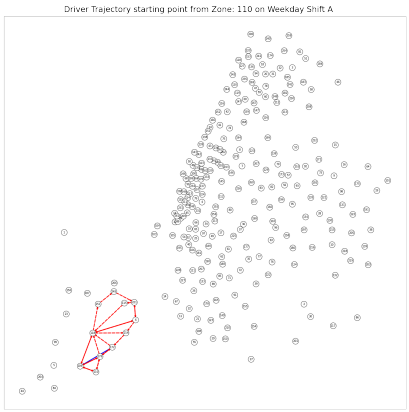

Fig. 21. Driver Trajectory for trips starting at Taxi Zone 110 (Staten Island)

Understanding our results and simulator behavior helped us conclude that because all the three policies we evaluated included repositioning to an adjacent neighboring zone only, our trip trajectory for areas like Staten Island could never move out of an never ending loop within itself. This boosts our confidence that the simulator we developed was able to capture this behavior of the policies we tested. In an ideal scenario, an optimal policy would reposition the driver to a zone with higher arrival rate and not confine the reposition only to adjacent zone.

E. Generalized Policy Iteration

Finally, at the conclusion of policy evaluation, it is possible to undertake policy improvement and to then follow a process of generalized policy iteration. The figure below from Sutton [6] illustrates this process.

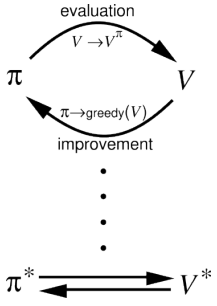

Fig. 22. Generalized policy iteration

For policy improvement, the next step would be to review the values and change the policy to a greedy approach. This step is theoretically straightforward, but given the spatiotemporal data, there would be additional considerations that would need to be taken into account.

For example, all of the policies listed above that were evaluated involved repositioning, but only to adjacent zones. Rather than a policy governed by a fairly short sighted metric (hot spot tracing based on the immediate gain in probability matching), one possibility during policy improvement would be to allow repositioning to more zones beyond just adjacent zones, but the benefits of traveling to a distant zone and obtaining a passenger quickly would need to be balanced with the costs of such a trip. In this example, given the temporal nature as each driver works a set 11-hour shift, repositioning to a far off-zone would likely incur heavy implicit opportunity costs even if the explicit marginal fuel cost is low.

The hope would be that the state values found by our policy evaluation process would be sufficient such that with the correct calculation of the distance, traveling time, and additional costs (e.g., tolls) for repositioning to non-neighboring zones, we could compare potential all potential repositioning actions by incrementing time by the travel time and costs by fuel costs passed on mileage and any tolls and considering the different evaluated future state values at each potential zone we could reposition to. However, state values are only calculated in 30 minute intervals which may be an overly large margin for repositioning policies.

One possible alternative approach could be to assign the driver an hourly cost, for example, using the amortized lease payments per driver of $9.58 to $11.58 per hour [5] as reported by the TLC or to more directly assign an average hourly wage and have the goal be to then make wages in excess. The approach decided on would need to be calibrated well by looking at simulation results, and simulation results would need to be repeatedly checked ensure that behavior is realistic.

Also, when undergoing an generalized policy iteration process, additional care would need to be taken to ensure that the model does not over-optimize.

VII. Miscellany

A. Potential Future Expansions of Research

There are many future potential avenues through which we believe this research could updated, extended, and expanded.

1) Updating and comparing: This research relied on data from June 2013 as individual Taxi behavior could be tracked to better estimate. Since June 2013, many large developments have occurred. One of the largest is the increasing popularity of mobile FHVs such as Uber and Lyft. One of the reasons for their popularity was their ability to attract previously censored demand. For example, whereas potential taxi passengers may have previously tried to hail a taxi, but decided to give up and take the subway after 15 minutes of fruitless efforts, this same passenger can now request a vehicle on his or her mobile device and have it come to his or her location within a specified time window.
   One possible extension of our research could be to focus almost exclusively on demand estimation and see whether more recent data, especially from FHV can validate or refine demand censoring observations. For reference, FHVs typically had limited reporting requirements, but as of 2019, High Volume FHVs that currently dispatch or plan to dispatch more than 10,000 FHV trips per day are subject to greater transparency in reporting[2].
   Any such efforts, however, would need to take into account that the rising popularity of FHVs captured not just previously uncaptured demand but a greater share of ridership volume from medallion taxis and may have created or benefited new demand as well. In addition, there have been multiple changes such as cost structure and rules (e.g., in 2015, fares were increased by an improvement surcharge) and macroeconomic changes and changes in customer sentiment (e.g., COVID) that would need to be considered for such an endeavor.
2) Multi-party model: As described above, one possibility could be to shift the simulator for policy evaluation to be based on a multi-party model. This would introduce not just additional computation complexity, e.g., for the simulator to simultaneously track thousands of taxis at once, but also challenges down the line for policy improvement, e.g., not all taxis may follow the exact same policy as it could lead to overcrowding so would need to balance supply and demand and assign policies given the current state and actions of other drivers.
3) Extending destination probabilities for times of censored demand: As described earlier, there is an opportunity to instill a mechanism by which if there is censored demand, the destination will be sourced more probabilistically rather than strictly from the Markov matrix of previous destinations demanded from that location. This would require the creation of a small model or the assumption of more direct coordinates and a map/directions API to estimate trip duration and fare.
4) Alternative approaches to policy evaluation: For this first approach, we chose to conduct policy evaluation through Monte Carlo evaluation, however there are several other algorithms and methods that could be employed including Monte Carlo Rollout and Dynamic Programming.
5) Carrying through Generalized Policy Iteration: Finally, generalized policy iteration could also be conducted, but with the reservations noted in the section above with respect to the optimization process and ensuring the model does not over-optimize.

B. Acknowledgments

We would like to express our gratitude to Vineet Goyal and Zhiwei Qin for guiding and mentoring us throughout the term. It has been an invaluable and thrilling experience trying to create a novel process for policy evaluation, and it has been a pleasure to work with them in creating this. We would also like to thank Sining Chen, for organizing the capstone course and providing us with advice and guidance along the way.

[2]https://www1.nyc.gov/assets/tlc/downloads/pdf/trip_record_user_guide.pdf

C. Contributions

1) Danyang Han: Data transformation, processing data for arrival rates by time zone, analyzing pick up/drop off counts, testing proposed models for Poisson based vacant taxi probability, data visualizations and geospatial heatmaps, Simulator and Environment Development

2) Anita Pinto: Cleaning and preprocessing data, data visualization, formulating and testing models for probability matching calculation, inter-pickup time arrivals and lost demand estimation, demand optimization, policy set up, evaluations, simulations, coordinating between teammembers.

3) Elizabeth Yum: Organizing materials for meetings, setting up logistics, managing progress, transition matrices, calculation tools for trip distance, duration, and fare, approach for supply, collating and documentation of the results and reports.

D. Ethical Considerations

Our proposed research method was that of creating a simulation based on historical data. As such, there were no research participants that could be impacted or harmed in any way. If this sort of research were extended to allow for repositioning recommendations to taxi drivers on a live, production basis, such a system would need to abide by rules of fairness, social responsibility, legal compliance, and public health & safety among others, but we did not believe that our approach marked a meaningful step toward such a sophisticated system to warrant investigating questions of fairness in deployment here. With regard to privacy concerns, the dataset we used was publicly accessible. It contained no personally identifying passenger information. It did contain taxi medallions which could be traced back to individual medallion owners, however, this was information that had been published online for widespread use by the NYC TLC at the time, and our usage of taxi medallion numbers was strictly to observe and model behavior by individual taxis in the aggregate.

E. Github

We have hosted our code on Github: https://github.com/eyy4/Spr-2021-DiDi-RL

REFERENCES

[1] "Improving Efficiency and Managing Growth in New York's For-Hire Vehicle Sector," New York City Taxi and Limousine Commission and Department of Transportation, June 2019, https://www1.nyc.gov/assets/tlc/downloads/pdf/fhv_congestion_study_report.pdf

[2] Mersereau, Adam J. "Demand estimation from censored observations with inventory record inaccuracy." Manufacturing & Service Operations Management, vol. 17, no. 3, 2015, p. 335, ISSN: 1523-4614, https://doi-org.ezproxy.cul.columbia.edu/10.1287/msom.2015.0520

[3] Yu, Xinlian, Gao, Song, Hu, Xianbiao, and Park, Hyoshin, "A Markov decision process approach to vacant taxi routing with e-hailing", Transportation Research Part B: Methodological, Volume 121, 2019, Pages 114-134, ISSN 0191-2615, https://doi.org/10.1016/j.trb.2018.12.013

[4] Tian Wang, Yingyu Cao, Bo Jumrustanasan, Tianyi Wang, Xue Xia, Vineet Goyal, Zhiwei Qin, Shuaiji Li, and Qun Li, "Reinforcement Learning for Taxi Driver Re-positioning Problem in NYC,' unpublished.

[5] "2014 taxicab fact book," New York City Taxi and Limousine Commission, Tech. Rep., 2014, https://www1.nyc.gov/assets/tlc/downloads/pdf/2014_tlc_factbook.pdf.

[6] Sutton, Richard S. and Barto, Andrew G. "Reinforcement learning: An Introduction," MIT Press, 2nd edition, 2015.