Reimagining Bus Routes with Q-Learning and an Equity Based Reward Model

Justin Tinker
jatinker@cs.stanford.edu

Elizabeth Zhu
elizhu@stanford.edu

Helena Aytenfisu
haytenf@stanford.edu

December 11, 2023

1 Abstract

Public transit is an essential service provided by the government that has wide-reaching social and economic impacts. For those who do not have reliable access to a vehicle, public transit is often the sole method by which to access essential services such as grocery stores, schools, hospitals, and other opportunities. This group of people are often referred to as 'dependent riders'. As of now, many bus routes are implemented in a way that perpetuates socio-economic, geographic, and racial inequities, and optimize for factors such as aggregate ridership rather than equitable access for everyone. This project explores how to integrate equity based metrics into bus route planning, and applies this method to improve existing bus routes in major cities in the United States, including San Francisco and Cleveland.

2 Introduction

Bus Network Design and Frequency Setting (BNDFS) is well observed as a sequential decision making process in urban planning literature [2]. Multiple approaches have been considered. A common framework involves ingesting a list of bus stops and passenger counts, then choosing which buses will connect which stops (routing) and how many buses to run along each route (frequency). Selecting one stop affects how we select subsequent stops, making this sequential. Numerous reward functions are also studied in the literature. While the classical minimization of aggregate travel time still dominates, new research is exploring more equity based objectives [2]. Current bus routes tend to be concentrated in densely populated urban regions, making it difficult for people from more remote, far-off areas to use the bus to access essential services. On top of this, bus routes are often designed in a way that results in large 'dead zones' where a lack of coverage makes it difficult to get to one's desired destination. In addition, bus routes do not prioritize 'dependent riders' over 'choice riders', meaning that those who rely on buses often have ironically longer commutes than their more privileged counterparts, who have access to alternate modes of transportation to reach their destination. Taking these factors into consideration, using publicly available data from bus routes in major cities, including San Francisco, New Orleans, and Miami, we model the problem of route planning as a sequential decision-making problem and apply Q-learning to optimize bus routes not only to minimize travel time, but to maximize equity.

3 Methods and Results

We explored two Q-Learning methods for optimizing bus routes in this paper:

1. In-place Modification uses Q-learning to identify small route modifications (i.e. adding a stop, changing route frequencies) to improve the overall bus route network.

2. Route Creation uses Q-learning to develop new routes that share the same origin and final destinations as the existing routes, driven by an equity-based reward function.

We begin with a simplified representation of the existing bus network. In our in-place modification approach, a bus network is represented by six components:

1. $\boldsymbol{R}$, the list of routes in the network.
2. $\boldsymbol{S}$, the list of stops in the network.
3. A mapping from each bus route r to an ordered list of stops $\boldsymbol{s}$ visited by r.
4. A mapping from each bus stop s to an unordered set of routes $\boldsymbol{r}$ that visit s
5. A mapping from each bus route r to a frequency f, where f is equal to the minimum interval between buses running along r.
6. A graph G where two stops s_a and s_b are connected by a directed edge e if there exists any route in R that travels directly from s_a to s_b

Each bus network is loaded from publicly available GTFS (Google Transit Feed Specification) data and converted into the representation described above (for an example, see Fig. 1 below).

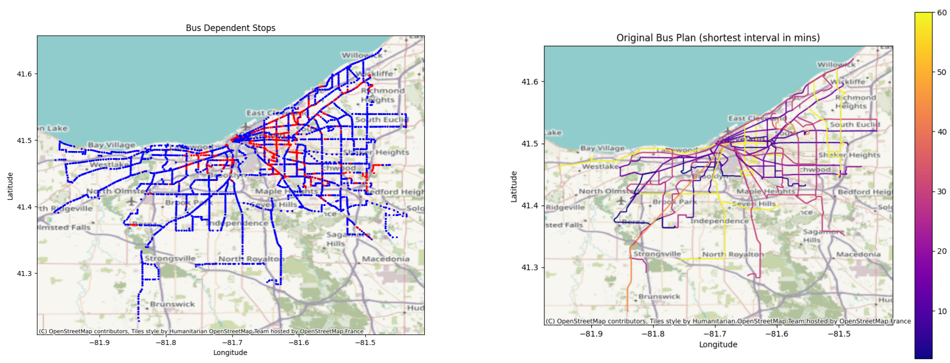


Figure 1: Cleveland Bus Network

For each stop, we add features from Census (ACS 2021) and Open Street Maps (OSM) data. A stop is defined to be near a point of interest (POI) if it is within 200 meters of the location coordinates in OSM. A stop is defined to be in a bus dependent area if

1. At least 40% of units in the census block group are rentor-occupied
2. Less than 50% of households own a car
3. At least 20% of households are within 200% of the federal poverty line

The following table (Table 1) shows an example table of stop features.

Table 1: Sample Stops Data

Stop ID	**Stop Name**	**Lat**	**Lon**	**Near Grocery**	**% Rented**
04759	FULTON RD / BUSH AV	41.47706	-81.708561	True	1.0
06263	MILES AV / E 123RD ST	41.445392	-81.596544	False	0.23

3.1 Method 1: In-place Modification

We define 7 benchmarks to evaluate a bus network. Each benchmark represents an individual traveling from an origin stop to either a single destination stop or any of a list of destination stops. The fastest route is simulated using a multisource Dijkstra's algorithm with the following parameters:

1. Penalty for transferring between buses: 5 minutes
2. Wait time per stop: 30 seconds
3. Driving time: Manhattan distance between stops divided by 30 miles per hr
4. If transfer between buses is required, the wait time equal to the frequency of the required bus route is added

The network's score is the average time it takes to complete each of the following benchmarks:

1. Randomly chosen bus-dependent stop to any hospital
2. Randomly chosen bus-dependent stop to any park
3. Randomly chosen bus-dependent stop to any bar
4. Randomly chosen bus-dependent stop to any place of worship
5. Randomly chosen bus-dependent stop to a randomly chosen Starbucks
6. Randomly chosen bus-dependent stop to a randomly chosen McDonalds
7. Randomly chosen bus-dependent stop to a randomly chosen bar

Notably, the latter three are designed to simulate stochastic work commutes. Furthermore, there are 5 possible actions, each with a parameter X:

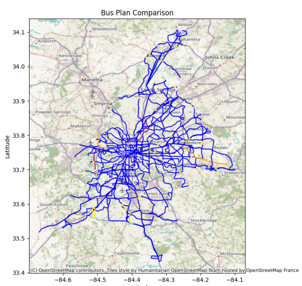
1. Increase a route's frequency (by X minutes)
2. Decrease a route's frequency (by X minutes)
3. Replace a random stop on a route(with a new stop within X meters)
4. Add a stop at a random position along on a route (new stop must be within X meters of the current stop at that position)
5. Remove a random stop from a route

To ensure that the improved bus networks do not require more resources or labor than the old bus network, we add a constraint. We define the total bus-minutes of a network to be the total amount of time drivers must spend to complete a single day's worth of routes. For example, if one route takes 30 minutes to complete, and runs 5x per day, the total bus minutes to complete that route are 150. We constrain the optimizer to identify networks that require more than 110% of the existing network's bus minutes. We use Q learning with e-greedy exploration to find the optimal choices of actions and parameters.

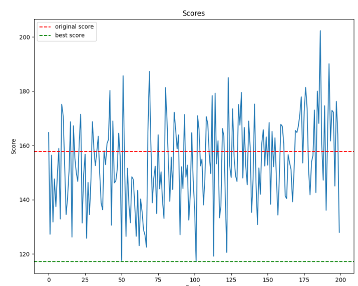
3.2 Results: In-Place Modification (Atlanta + San Francisco)

In Atlanta, we decrease the average time required to complete a benchmark route from 158 minutes to 117 minutes (-35%), with a 0.4% increase in total bus minutes.
In San Francisco, we decrease the average time required to complete a benchmark route from 116 minutes to 81 minutes (-43%), with a 0.2% increase in total bus minutes.
In Cleveland, we decrease the average time required to complete a benchmark route from 128 minutes to 86 minutes (-32%), with a 9.5% increase in total bus minutes.
In New Orleans, we decrease the average time required to complete a benchmark route from 87 minutes to 63 minutes (-38%), with a 1.3% increase in total bus minutes.
In Miami, we decrease the average time required to complete a benchmark route from 97 minutes to 61 minutes (-59%), with a 2.1% increase in total bus minutes.

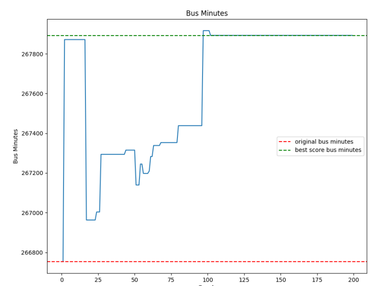
In Figures 2 and 3, blue lines represent unchanged routes, yellow lines represent additions to the route network, red lines represent subtractions, orange lines represent routes with increased frequencies, and purple lines represent those with decrease frequencies.



Atlanta Bus Plan Comparison

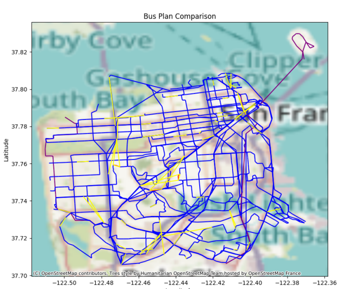


Atlanta Scores

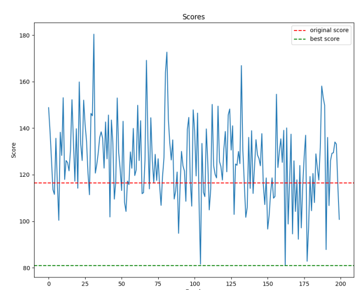


Atlanta Bus Minutes

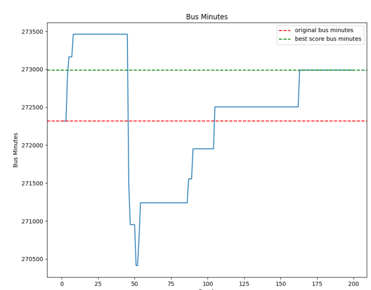
Figure 2: Atlanta In-Place Modification Results



Difference Map



Scores



Bus Minutes

Figure 3: San Francisco In-Place Modification Results

3.3 Method 2: Route Creation using Q-Learning

In our route creation approach, we represent our bus route planning problem as a Markov decision problem. Our approach takes in an existing route, and construct a new, more equitable route with the same start and end stops. We sequentially build a route from the starting bus stop, using Q-Learning and a reward function motivated by an emphasis on equitable access. We used the following states, actions, and exploration strategy:

State: The current route $[s_1, s_2, s_3, ..., s_i]$ up to, and including, the present stop.
Action: Add stop s_i to route r, from an array of nearby stops.*
Exploration Strategy: Epsilon-Greedy (with gradual decay of epsilon)
Q-Learning Update: $Q(s,a) \leftarrow Q(s,a) + \alpha \cdot (r + \gamma \max_a Q(s',a) - Q(s,a))$

*Note: The array of nearby stops to choose from was restricted to fall within a 500m to 1000m radius of the current stop, in addition to being restricted to nearby stops that were closer in distance to the end stop than the current stop, as a constraint for the action space.

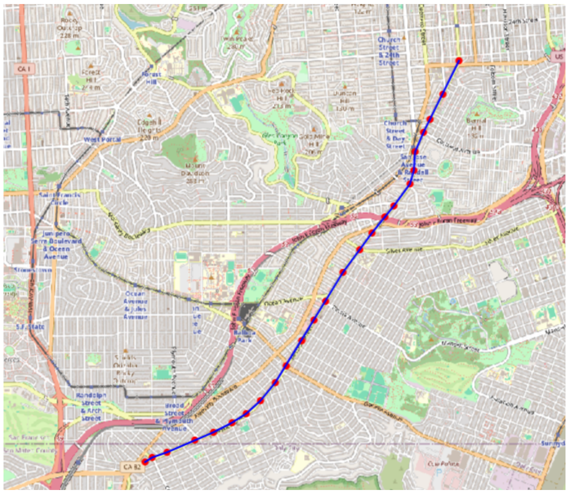
Reward Function: Our reward r was defined to explicitly integrate equity based rewards, taking into consideration various best practices in equity-driven transit planning and encoding them into quantitative rewards. A few standard equity indicators in transit planning are access to employment, schools, healthy food, and medical facilities [1].

Specifically, r was applied to a state-action pairing based on the following 3 criteria:

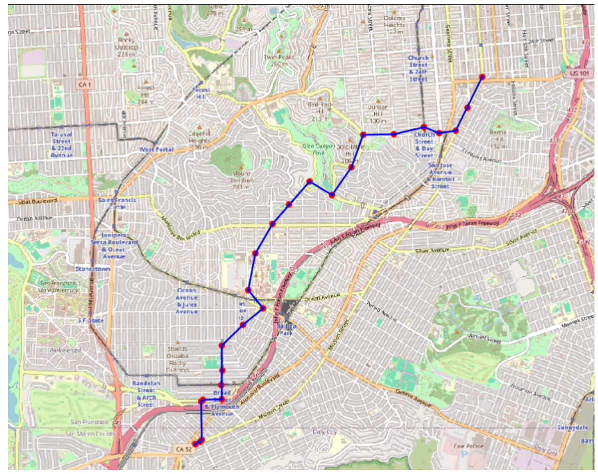
(1) whether the route brings bus service access to a bus dependent area
(2) whether the route brings the rider access to economic opportunity (jobs), evaluated by a proxy measure of having a Mcdonalds or Starbucks near at least one stop on the route
(3) whether the route allows the rider to access essential services, evaluated by determining if a hospital or grocery store is accessible from at least one stop on the route.

The state is evaluated based on each criterion independently. A large reward is applied for each of the three criteria that the action satisfies if the previous route did not previously satisfy these conditions. For routes by which a criterion was already satisfied, any action that results in a state where the conditions continue to be satisfied receives a smaller reward. No meaningful reward is applied to a state action combination by which the conditions of a criterion are not met (a baseline reward is applied to all states to prevent negative values). Lastly, our reward factored in distance to the end stop, incentivizing our algorithm to choose next stops closer to the end destination by subtracting this distance from the reward.

3.4 Results: Route Creation Q-Learning (San Francisco)

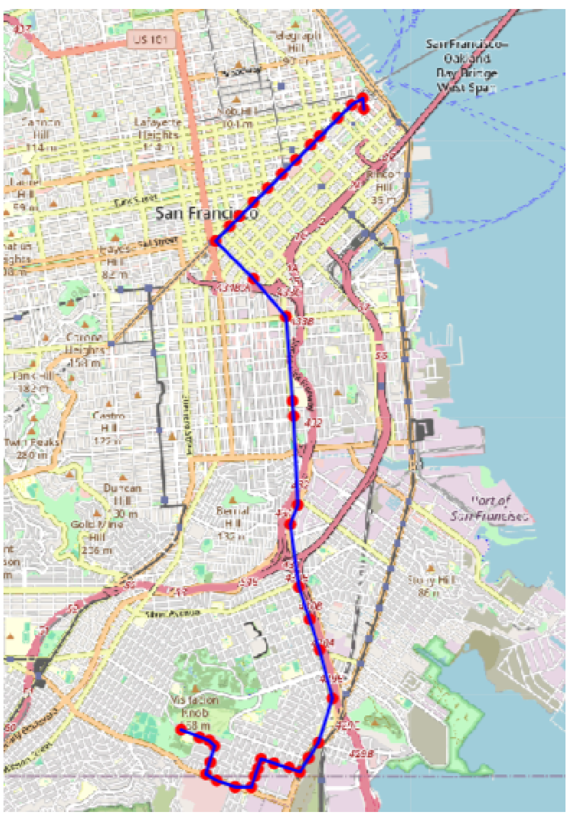


Original SF Muni Route 14

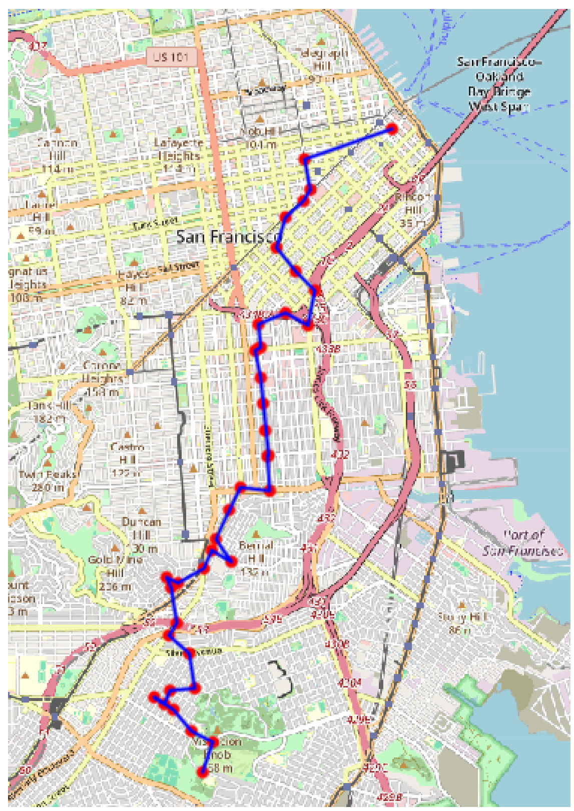


Optimized SF Muni Route 14

Figure 4: Mission (14): Pre and Post Q-Learning



Original SF Muni Route 9R



Optimized SF Muni Route 9R

Figure 5: San Bruno Rapid (9R): Pre and Post Q-Learning

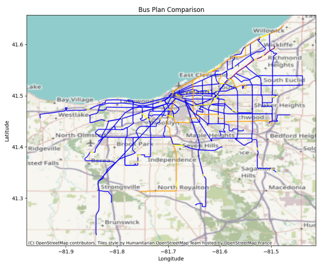
4 Conclusion

In this paper we have explored two different approaches to optimizing bus routes in major cities using Q-Learning, both motivated by an equity-oriented reward function. For our first approach, we modified routes in place using Q-Learning to optimize adding stops and changing frequencies of existing routes. For our second approach, we created routes with the same start and end as existing routes using Q-Learning to generate an optimal route.

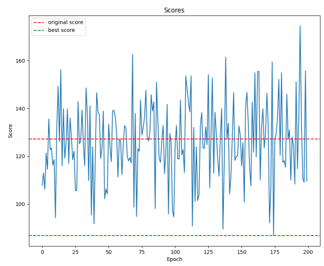
In contrast to traditional bus route optimization methods, we prioritized equity-based metrics in determining our modified routes. For instance, our model considered bus stops that would make it easier for a given passenger to travel to essential destinations, such as hospitals and grocery stores, catering towards the needs of the riders who depend on buses the most. As a result, routes were not chosen simply with the goal of routing efficiency in mind, like typical approaches; although distance and time were certainly factors, our reward model would also incentivize detours that would improve overall accessibility for passengers.

5 Appendix

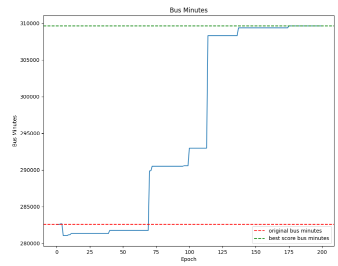
5.1 In Place Modification Results for Various Cities



Difference Map

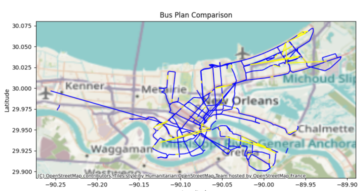


Scores

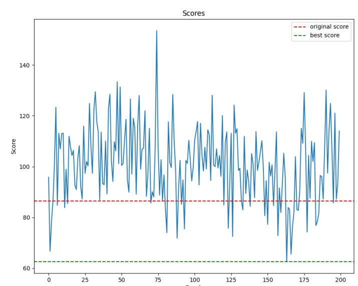


Bus Minutes

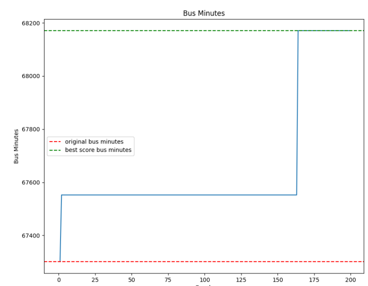
Figure 6: Cleveland Results



Difference Map

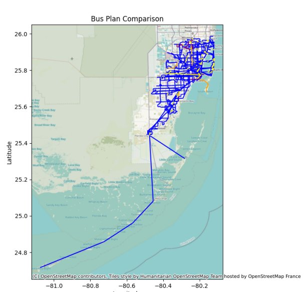


Scores

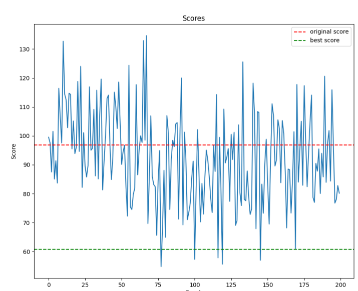


Bus Minutes

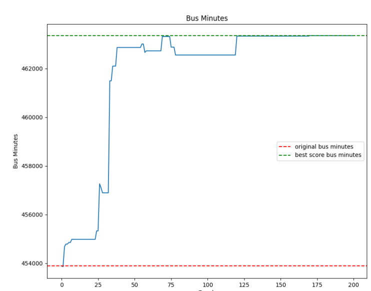
Figure 7: New Orleans Results



Difference Map



Scores



Bus Minutes

Figure 8: Miami Results

References

[1] Williams; Boyd; Keita; Kramer. Transportation equity needs assessment toolkit. Technical report, University of South Florida, University of Texas at Arlington, 2021.

[2] Sunhyung Yoo and Jinwoo (Brian) Lee. Revising bus routes to improve access for the transport disadvantaged: A reinforcement learning approach. *Journal of Public Transportation*, 25:100041, 2023.

6 Individual Contributions

Elizabeth: Q-learning (Method 2)
Helena: Reward model/functions (Method 2)
Justin: Data processing Modeling, Q-learning (Method 1)