# Genetic Algorithm for Equitable Neighborhood Service

AI4SG

Jacob Emmerson
Chris Hinson

July 30, 2024

# Outline

# Target Problem

We seek to leverage DNNs and a genetic algorithm to detect equity disparities in the Pittsburgh public transit system, and identify a single route change which will have the greatest impact in minimizing total disparity in the system.

# Equity Problem

There have been several separate investigations into how equitable pittsburgh's current public transit system is. These reports look at a wide variety of indicator variables to form a single opaque equity score, typically on the granularity of a census tract. We sought to build upon and extend this work, by taking both their indicator variables along with our own measurements, to detect equity disparities in the system.

# Our Contributions

We aim for two primary contributions:

1) Equity-focused demand predictions for currently unobserved – but potential – bus segments.
2) Midpoint mutations when modifying transit routes for faster solution convergence.

# Datasets

Almost all dataset are held in memory as numpy arrays with a strict ndtype. For data entries with fields that represent an entry in another data blob, raw indexes are used. This setup means that all of these datasets must be held in a single container class so they may access dependency datasets.

| **spatial** | **data blob** |
|---|---|
| Allegheny county census tracts | transport data by census tracts |
| Allegheny county zoning zones | PRT equity index of mobility need |
| Bus Stop reach-ability network | google maps Places API |
| | PRT stops |
| | PRT segments by route |

# Data Structures

**GAENS class**

- Holder for datasets and Genetic Algorithm

**Population**

- Poluation: Collection of Individuals
- Individual: Collection of Genes
- Gene: Route
- Route: Collection of route segments
- Route segment: Source and Sink Bus stop, along with usage information

# Consumer Demand

Most literature on TRNDP seeks to maximize the demand coverage within a network. Our objective is to maximize the equitable demand in a network by redistributing predicted demand into under served neighborhoods.

We have ground truth demand data for our current route segments, but how can we make predictions on current unobserved routes? We can train a model to make demand predictions utilizing locally available information between two bus stops.

# Consumer Demand *continued...*

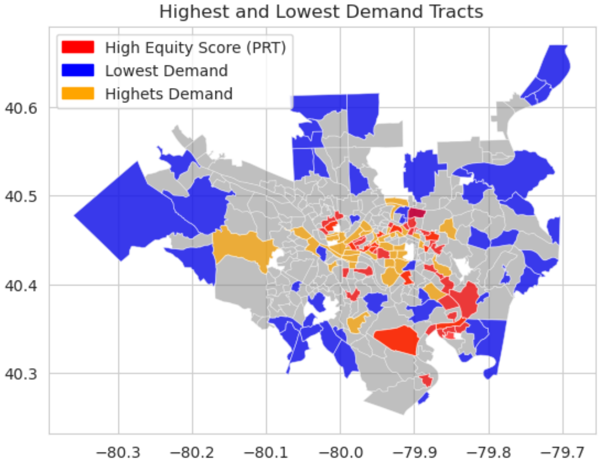



Figure: Equity and Demand by Census Tract

# Neural Network

To make demand predictions on the edges, we utilize local features to each node and train a neural network to learn a function $w : (v_i, v_j) \to \mathbb{R}^+$.

- 408 input features – 204 features for each node
- 4 hidden layers w/ ReLU activation
- Batch normalization
- p(Dropout) = 0.25

# Evolutionary Algorithms

Prior research on the TRNDP has found brute force solutions to be inefficient; thus, evolutionary algorithms have been proposed for optimization over transit networks [4], [7], [9], [1], [10], [6], [5].

In our approach, we utilize a *genetic algorithm* to optimize over a set of bus lines. Genetic algorithms are a type of meta-heruistic algorithms can find near-optimal solutions over a set of configurations by constraining the search space to a reasonable set of decision variables.

# Evolutionary Algorithms *continued...*

Let $G = (V, E)$ denote a graph with a set of edges $E$ and vertices $V$.

- An edge $e_{ij} \in E$ has a weight equal to the predicted demand between $v_i$ and $v_j$ for $v \in V$.

We want to find a set of bus lines which maximizes demand coverage in the network while accounting for equity factors.

# Populations

Let $P$ denote a population of individuals $p \in P$ where $|P| = n_{pop}$. Each $p$ has a representation of the transit network called a chromosome consisting of $L$ genes.

An individual is a set of genes $p = \{g_1, \ldots, g_L\}$ where $g_l = \{v_1 \ldots v_i\}$ for a bus line of $i$-stops.

- $g_1$ : initial bus stop
- $g_i$ : final bus stop

# Populations *continued...*

Let the initial population be denoted by $P_0$ which consists of the current PRT bus network.

Within $n_{gen}$ generations, we want to find the individual which minimizes the total edge weights over the graph. Let our solution be $p^*$ where

$$p^* = \arg\max_{p \in P} f(p)$$

and $f$ is a function denoting the total utility of a given individual.

# Tournament Selection

We use *tournament selection* to determine which individuals will be used for crossover.

From a random sample of $k$ individuals, the two with the highest fitness become the parents for two new individuals in the population which is created using uniform crossover.

# Uniform Crossover

In genetic algorithms, crossover is used to combine two high quality solutions into a potentially higher quality solution.

With the two parents selected, each gene in their chromosome has a probability of being swapped according to a crossover parameter $\xi_c \in [0, 1]$.

# Elitism

Since crossover does not always created a better quality solution, to ensure that we are always finding better solutions we can use elitism.

A small proportion of the most-fit individuals determined by $\phi \in [0, 1]$ is passed into the next generation without any modifications.

- Tournament selection is repeated $(n_{pop}(1 - \phi))/2$ times to maintain a constant population size.

# Mutations

In each individual selected for mutation, the gene with the lowest fitness is mutated according to a probability $\xi_{re} > 0.5$.

- **Small Mutation**: a more frequent mutation with a probability $\xi_{re}$. One of the terminal nodes is randomly selected and deleted according to a probability $\xi_d$. If the node is not deleted, $(1 - \xi_d)$, a new terminal node is added by randomly selecting an adjacent node.
- **Big Mutation**: a less frequent mutation with a probability $1 - \xi_{re}$. Similarly, one of the terminal nodes is selected at random but guaranteed to expand by greedily selecting the new terminal node.

# Mutations *continued*...

**Midpoint Mutation**: a variation of the **small mutation** with a probability of $\xi_{re}$. Rather than relying exclusively on terminal nodes, we randomly select a non-terminal node according to a uniform probability distribution.

- Motivation: the current terminal nodes of some bus lines are set due to nearby facilities or lot to swap driver, thus a change in these nodes may not be truly beneficial.

With the new non-terminal node selected, a nearby node is randomly selected and reconnected according to our network graph.

# Fitness

To determine which individuals have a higher utility, we evaluate each $p \in P$ at each generation utilizing a fitness function $f : p \to \mathbb{R}^+$.

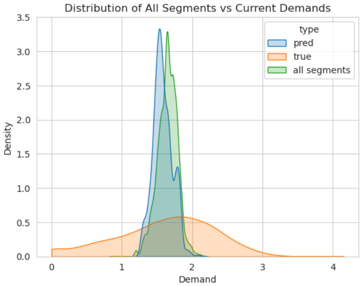
$$f(p) = \sum_{l=0}^{L} \sum_{r=0}^{R-1} w(v_r, v_{r+1})$$

Intuitively, our fitness function returns the sum of all equity-encoded demand predictions for every bus line in a network.

# Demand Model

| | MSE | MAE |
|---|---|---|
| Training | 0.4718 | 0.5512 |
| Test | 0.4774 | 0.5542 |

Table: Demand Error



# Next Week's Outline

i. Increase the robustness of our algorithms to handle data-driven errors
ii. Pilot program to test changes during off hours
iii. Improved feasibility constraints on route mutations, using road network path finding and real world traffic information
iv. Finish demand model fine tuning

# References

[1] B. Bulut, M. Günay, K. Özgün, and J. Ledet, “Optimizing bus lines using genetic algorithm for public transportation,” *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVI-4/W5-2021, pp. 131–136, Dec. 2021. DOI: 10.5194/isprs-archives-XLVI-4-W5-2021-131-2021.

[2] T. R. Dillahunt and T. C. Veinot, “Getting there: Barriers and facilitators to transportation access in underserved communities,” *ACM Trans. Comput.-Hum. Interact.*, vol. 25, no. 5, Oct. 2018, ISSN: 1073-0516. DOI: 10.1145/3233985. [Online]. Available: https://doi.org/10.1145/3233985.

[3] L. Fan and C. Mumford, “A metaheuristic approach to the urban transit routing problem,” *J. Heuristics*, vol. 16, pp. 353–372, Jun. 2010. DOI: 10.1007/s10732-008-9089-8.