Research Project Proposal
CSCI 633: Bio-Inspired Intelligent Systems
(Spring 2019)

James Le and Michael Peechatt

03/29/2019

1 Motivation

For our project, we are interested in solving *the travel time optimization problem for taxi vehicles*. This problem is very relevant given the rapid and significant evolution that is happening in urban transportation. The emergence of Internet-connected smartphones has allowed us to plan and optimize our daily commute. Today, ride-sharing companies like Uber and Lyft are using such technologies to gather large amounts of data and improve the efficiency of transportation systems, laying the ground for a more connected and centrally controlled transportation structure.

This problem is related to *vehicle routing* - the optimization of each vehicle actions to maximize the system efficiency and throughput. By optimizing the time travel, we can decide which vehicle to assign to each ride request. The vehicle routing problem is a combinatorial optimization problem where the number of feasible solutions for the problem increases exponentially with the number of customers to be serviced. As such, it is relevant to our class in a sense that we can use evolutionary / nature-inspired algorithms to generate possible solutions.

2 Dataset

The data that we will work with is the NYC Taxi and Limousine Commission Trip Record data, which can be accessed at this URL: *https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page*. We have the monthly data within each year from 2009 to 2018 for Yellow Taxi, Green Taxi, and For-Hire Vehicle.

For the Yellow and Green Taxi, the trip records include fields capturing pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts.

For the For-Hire Vehicle, the trip records include fields capturing the dispatching base license number and the pick-up date, time, and taxi zone location ID.

3 Experiments

We start with taxi data, use this data to make predictions for travel times between locations, and then run nature-inspired algorithms to optimize the total travel time. Here are the experiments we will conduct:

- *Explore the data and do feature engineering*. Since we are given each location's coordinates, let's calculate the Manhattan distances between each pair of points and count the longitude and latitude differences to get a sense of direction (East to West, North to South). Since we will be working with geospatial data, we plan to use a visualization tool like Tableau to visualize the data.
- *Pick a machine learning model*. We plan to experiment with different machine learning models from simple to complex, including linear regression, boosted trees, neural networks to predict the travel time - how long it takes to get from one point to another for each pair of points.
- *Optimize with nature-inspired algorithm*. We plan to experiment with different nature-inspired algorithms including genetic and ant colony to optimize the travel time.
- *Evaluate the results*. We will use mean absolute error as the evaluation metric for the prediction model. We will visualize our solution to the optimization problem with a demo.

4 Timeline

- *Week 11* (April 1st - April 5th): Pre-process, analyze, visualize, and engineer features for the NYC taxi dataset to get a broad overview of the most important features.
- *Week 12* (April 8th - April 12th): Try out different machine learning models (linear regression, XGBoost, neural nets) to do prediction.
- *Week 13* (April 15th - April 19th): Try out different nature-inspired algorithms (genetic and ant colony) to solve optimization.
- *Week 14* (April 22nd - April 26th): Evaluate the results and make proper changes to the models/algorithms if necessary.
- *Week 15* (April 29th - May 3rd): Prepare for the final presentation and write the final report.