Machine Learning for Cities Literature Review

Part I: Resources

*Fare and Duration Prediction: A Study of New York City Taxi Rides,* Chris Antoniades et al, link
*Artificial Neural Networks Applied to Taxi Destination Prediction,* Alex de Brébisson et al, link
*Predicting Taxi Utilization Information,* Stephen Andrew Atlas, link
*The Competitive Effects of the Sharing Economy,* Scott Wallsten, *link*
*Predicting Taxi Pickups in New York City,* John Grinberg, link

Part II: Paper Summarization

Title: *Fare and Duration Prediction: A Study of New York City Taxi Rides*
Authors: Chris Antoniades et al
Link: link
Abstract:

There were roughly 170 million taxi rides in New York City in 2013. Exploiting an understanding of taxi supply and demand could increase the efficiency of the city's taxi system. In this paper, we present a few different models to predict the number of taxi pickups that will occur at a specific time and location in New York City; these predictions could inform taxi dispatchers and drivers on where to position their taxis. We implemented and evaluated three different regression models: linear least-squares regression, support vector regression, and decision tree regression. Experimenting with various feature sets and performing grid-search to tune hyperparameters, we were able to achieve positive results. Our best-performing model, decision tree regression, achieved a root-mean-square deviation of 33.5 and coefficient of determination ($R^2$) of 0.99, a significant improvement over our baseline model's root-mean-square-deviation of 146 and $R^2$ of 0.73.

Summary:

Chris et al used several models to predict both the duration and the fare of taxi trips in NYC. Using all 170 million taxi trips in NYC in 2013, they started with the baseline model that achieves a RMSD of 145.78 and $R^2$ of 0.7318. They find out that the Decision Tree Regression model on feature set #1 performs the best, with RMSD of 33.47 and a $R^2$ of 0.9858. The feature sets they're using is listed below:

Feature Set #1: Zone, HourOfDay, DayOfWeek, HourlyRainfall
Feature Set #2: Zone, HourOfDay, DayOfWeek, Zone_HourOfDay, Zone_DayOfWeek_HourOfDay
Feature Set #3: Zone, HourOfDay, DayOfWeek, Zone_HourOfDay, Zone_DayOfWeek_HourOfDayl Zone_DayOfWeek_HourlyRainFall

To achieve a better prediction results, they proposed several methods for future studies, including predictions for arbitrary zones and time intervals, the implementation of a neural network regression, and the clustering feature template.