

# MIT Big Data Challenge: Transportation in the City of Boston

## Model of Prediction Challenge

Runmin Xu

Weighted linear regression is the main algorithm that I used to predict taxi demands. A set of potential models was generated at the beginning, from where each testing point would select the best-fit one automatically. The weight of a training point was affected by its similarity (determined by neighbor areas, time, weather and so on) with the testing point. Datasets that have been used are taxi pick-up, taxi drop-off, MBTA and weather, while twitter and events data act little.

Besides, I developed anomaly detection to keep track of the bias/variance of the models and to some extent avoid large errors, which is helpful for high demand spikes.

Useful features for city planners:

- Finding similarities between previous and current scenario
- Anomaly detection can be used to study special cases
- Regression model does work without future information