Road Traffic Prediction using IoT platform

Ish Kumar Jain

New York University

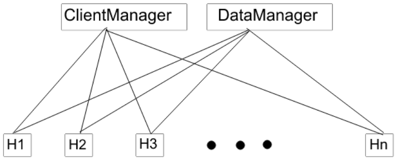
Spring 2017

About

- **Objective:** To evaluate a system that performs traffic map inference.
- **System Components:**
  1. **Mobile devices**- placed on every vehicle, collect data and report to the server.
  2. **Backend Server**- A controller that collect the data, send control messages to client (*e.g.* start/stop sensing) and performs traffic prediction and inference.
- **Source of Data:** Location, Accelerometer *etc*
- **System evaluation:** Mininet Simulation

Mininet Simulation

- Courtesy to Michael Corso and Ben Cullaj
- There is a Controller which opens two connections with every client— a DataManager and a ClientManager.
- The ClientManager is for exchanging control messages like start/stop sensing, sensing rates, select a sensor etc.
- The DataManager actually collects the data, analyse and store it in a database (txt file for now).
- It also does round-trip-time calculation and network analysis.



Road MAP

BridgeSegments

Bridge145
All items

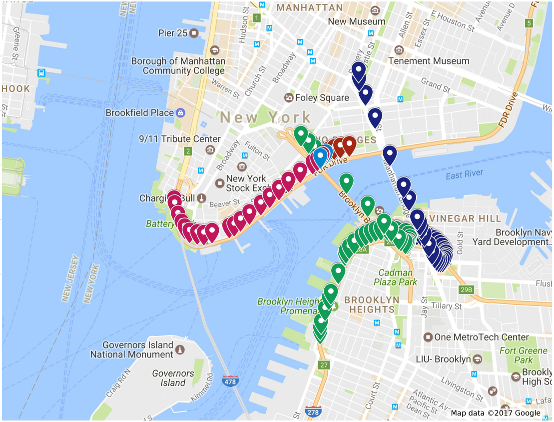
Bridge149
All items

Bridge150
All items

Bridge223
All items

Bridge443
All items

Manhattan and Brooklyn Bridge segments traffic data



Road MAP

BridgeSegments

Bridge145

All items

Bridge149

All items

Bridge150

All items

Bridge223

All items

Bridge443

All items

Man119

All items

Man122

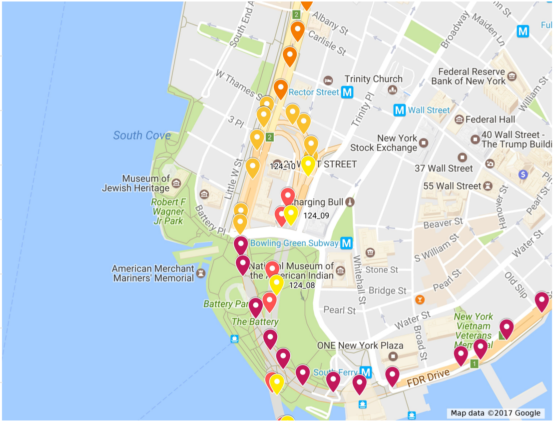
All items

Man123

All items

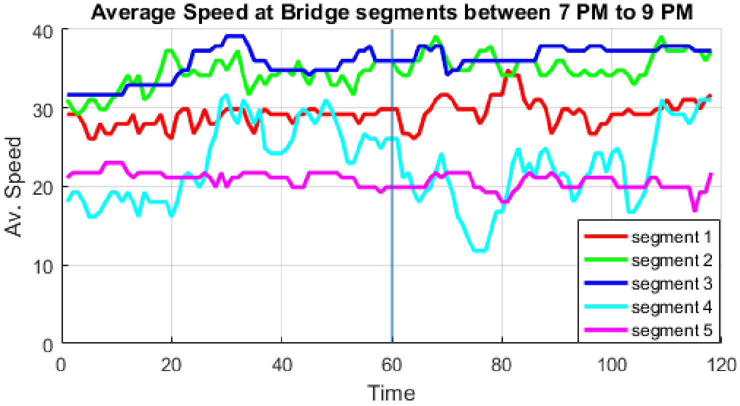
Man124

All items

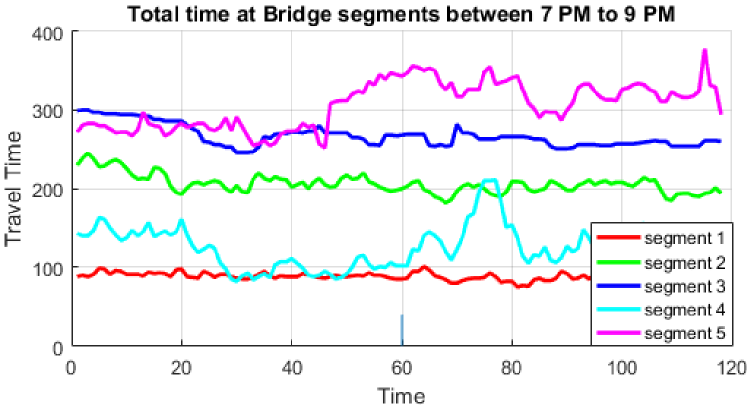


Manhattan and Brooklyn Bridge segments traffic data

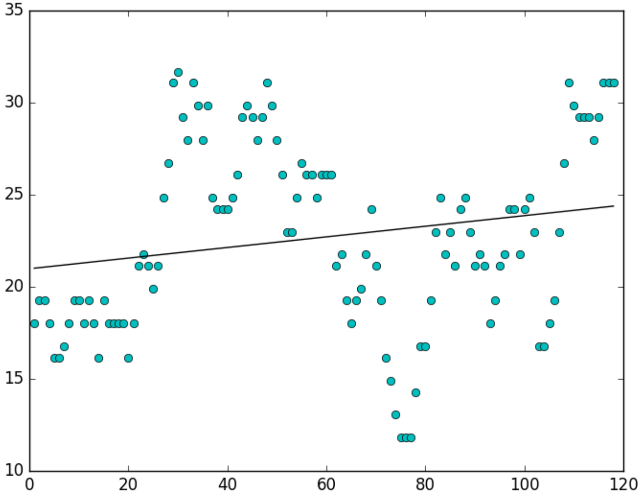
Time Series Data



Time Series Data



Time Series Data



Logistic Regression

- A classification problem.
- Quantized the speeds to many bins of size roughly 1 mile/hour.
- Trained on the data of last 5 speed samples to predict the next speed sample.
- Future traffic correlated with the past traffic.
- Error: Percentage of misclassified points.
- 2 hour data. 114 examples (80%training and 20%testing)
- Logistic regression framework:
  - softmax activation function
  - Objective: negative log likelihood (NLL) function
  - Optimization method: SGD
    - 1000 iterations
    - learning rate: 0.13

Logistic Regression

Segment#	Train Loss	Test Loss
1	63%	65%
2	84%	82%
3	73%	86%
4	84%	91%
5	65%	69%

Table: Classification error on bridge data for 5 segments

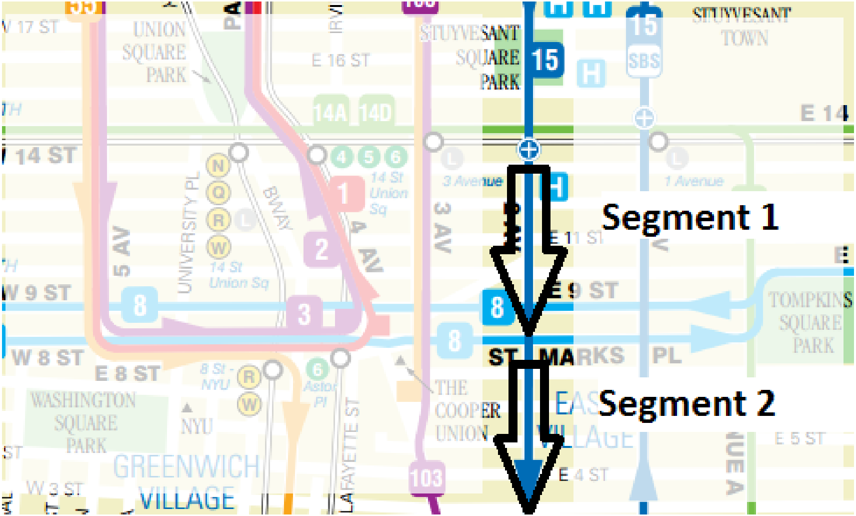
Logistic Regression

- Vary the total size of the dataset

DataSize	20	40	60	80	100
Train Loss	19%	56%	37%	44%	30%
Test Loss	50%	25%	50%	38%	55%

Table: Analysing the effect of varying dataset

MTA bus M15, 2nd ave- Segment Map



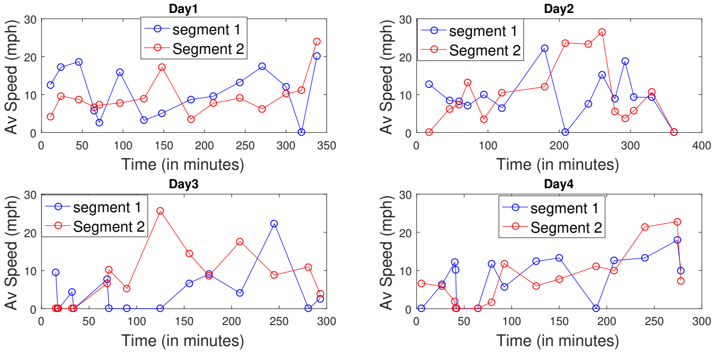
Data Collection

Procedure for Data Collection and Refinement.

- Segments are bounded by a rectangle. (Manually using Google Map). M15 Bus is located in both segments.
- For each segment-
  1. Bus Location & Time is extracted.
  2. Speed is calculated from consecutive locations.
  3. Average is done for speed values.
  4. Linear Interpolation is done when data is not available.
- Problem in synchronization with other bus data
  1. Bus Arrival times- different and sparse
  2. Frequency of arrival is different

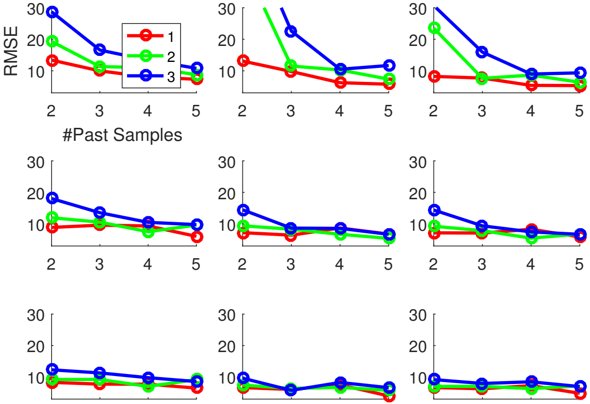
Time Series of Av. Speed Values on 2 Segments

- 4 Working Days. 5 hour data from 4PM to 9PM.
- Sampled at around 20 min (Bus frequency)



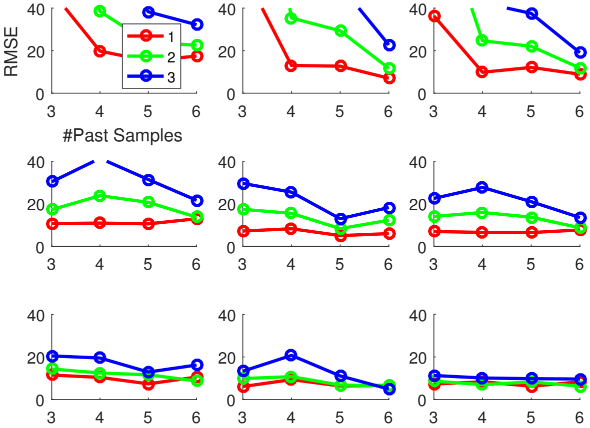
Linear Regression (On Segment 2)

1. Columns: 1. Seg2 only, 2. Seg1 only, 3. Both Seg1 and Seg2
2. Rows: 1. Day1 only, 2. Day1+Day2, 3. Day1+Day2+Day3
   - Red: Predicting future 1st sample, Green: 2nd and Blue: 3rd sample



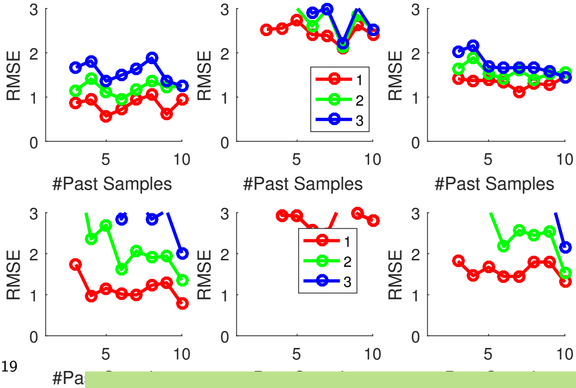
Quardratic Regression (Segment 2)

1. Columns: 1. Seg2 only, 2. Seg1 only, 3. Both Seg1 and Seg2
2. Rows: 1. Day1 only, 2. Day1+Day2, 3. Day1+Day2+Day3



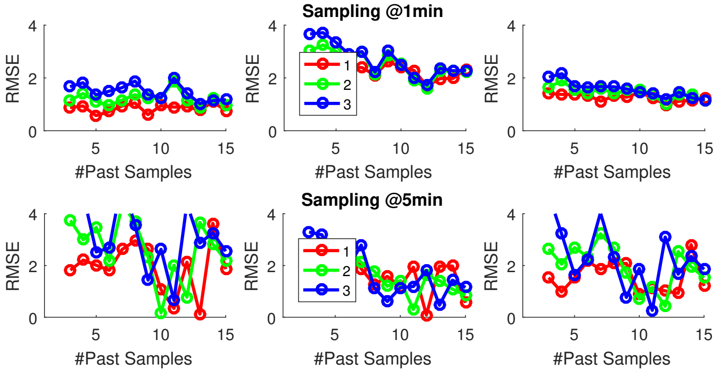
Back to Bridge Data (Predicting Manhattan Bridge Traffic)

- Columns: Training using 1. Man Bridge, 2. Brooklyn Bridge, 3. Both Bridge
- Rows: 1. Linear Regression, 2. Quadratic Regression



Sampling of Training Data

1. Bridge data: Same columns.
2. Rows: 1. Sampling at 1min, 2. Sampling at 5min



Future Work

- Real-time traffic prediction on MTA Bus Data
- Advanced ML Techniques:
  1. Generalized Linear Models. Non-linear models.
  2. Belief Propagation (Expectation Maximization)
  3. Bayesian Nets (Graphical Models)
  4. Arima Models (auto regressive time series model)
  5. Use Density instead of Speed.
- Network performance evaluation on Mininet.
- Data Collection using Mobile Phone (Android App- courtesy to Soumie)
- Thanks to Shiva R Iyer and Prof. Lakshmi Subramanian.