Flow & Travel Time Prediction using Machine Learning Algorithms to Optimize Traffic Responsive Coordination Plans: Santa Clara County Expressways

Shayan Khoshmagham, Data Scientist
Performance Analytics, Iteris

Ananth Prasad, Senior Civil Engineer
Santa Clara County Roads and Airports Department
Project Background

• Goals
  – Continually monitor and visualize performance along the County’s Expressway System
  – Improve the quality of data used in traffic responsive control
  – Extend traffic responsive control to use predictive flow and travel time data

• Timelines
  – iPeMS implemented in 2011
  – Predicted traffic flows and travel times in 2015-2016 (under MTC’s Next Generation Arterial Operations Program)
Santa Clara County Expressways

- 62 miles of expressways
- Trafficware ATMS.now with traffic responsive operations
- Mixture of loops and video detectors
- Bluetooth readers installed in 2015 for travel time/congestion monitoring
Canonical Detector Placement of Arterial

System detector. Sometimes measures speed.

Stop bar detector. Typically doesn’t measure flow.

Release detector

Advance detector
System Architecture

Field
- Video Sensors
- Traffic Signals, Sensors

SCC Facilities
- RTT Systems
- Naztec
- SCC FTP Server
- RTT-based data
- Naztec-based data

Iteris Hosting
- FTP pull
- FTP Server
- iPeMs data
- iPeMs
To the Cloud!

AWS Monitoring Console
Data

• 140 monitored intersections
  – 2,800 loop detectors
  – 2,000 video detectors
  – Cycle lengths and phase durations

• Collected from two separate systems

• Volume and occupancy/speed every 3 minutes
Data Quality
Data Processing

• Purpose
  – Send cleaned, filled-in volume and occupancy data back to the ATMS in real-time for use in traffic responsive operations
  – Use validated data in performance measures
  – Provide information on when and where equipment or communications are broken

• Imputation
  – Fill in with historical data
  – Send back data from “better” sensor
Arterial Imputation – Challenge

- Stations (either Naztec loops or RTT cameras) are treated completely independent of each other:
  - If loop is bad, impute using Time of Day median for that detector (otherwise, global median)
  - If camera is bad, we impute using Time of Day median for that sensor (otherwise, global median)

- Pick the “best” one of loops or cameras
  - More Observed data than Imputed data
  - More detectors (more # lane points)
  - More speed points
  - More occupancy points

<table>
<thead>
<tr>
<th>Freeway Imputation</th>
<th>Arterial Imputation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imputation is its own class</td>
<td>Subclass of arterials::summarizer</td>
</tr>
<tr>
<td>District ID</td>
<td>Feed ID</td>
</tr>
<tr>
<td>Different imputations for different station types</td>
<td>N/A</td>
</tr>
<tr>
<td>4 different methods (i.e. Local Neighbors, Global</td>
<td>2 methods (i.e. Temporal Median, Global</td>
</tr>
<tr>
<td>Neighbors, Temporal Median, Cluster Median)</td>
<td>Median)</td>
</tr>
</tbody>
</table>
Performance Measures

- Performance measures are pre-computed for every 5 minute period so they can be queried quickly.
- HCM-based calculation use detector and signal data to compute by movement:
  - Factored flow
  - Effective Green
  - Capacity
  - V/C ratio
  - Delay (uniform, incremental, control, total)
  - Queue length
  - Level of Service
Reporting

- Time Series
- Time of Day
- Day of Week
Traffic Responsive Operations

• Part of MTC’s Next Generation Arterial Operations Program 2015

• Implementation
  – 15-minute ahead predictions generated from cleaned data
  – Machine learning approach using K Nearest Neighbors Algorithm
  – Algorithm tuned to minimize mean absolute error while being fast and not requiring frequent retraining
Flow and Travel Time Prediction

- Machine learning algorithm, K Nearest Neighbor (kNN)
- Sensitivity Analysis: considered k=1, 3, 5, and 7; N=7, 14, and 21
- Notes: cannot have k larger than number of days in the comparison period
- Result: Result varies depending on # of days in comparison period
  - Comparison period = 14 days (2-weeks) and k=3 is optimal
  - (Comparison period = 21 days and k = 5 is also acceptable, but computationally more expensive)
- On September 12, 2016 at 7:14pm the deployed algorithm was tweaked to use N=14 and k=3.
Prediction Results

- Overall Mean Absolute Error (MAE): ~8 vehicles
- Blue Dots: *Observed* Flow
- Red Dots: *Predicted* Flow
- Gray Dots: *Historical* Flow
Travel Time Data Integration

- Bluetooth readers installed at strategic locations on 8 expressways
- Predict travel times provided back to SCC
- Performance reports for Bluetooth travel times and O-Ds implemented in iPeMS
Conclusions

• Higher-resolution, better data and web-based analysis tools are transforming the way agencies and consulting firms approach signal operations and maintenance.

• Trained algorithm was implemented and continues to run in real-time to generate 15-minute ahead flow and travel time predictions.

• The predicted values are collected by the County’s ATMS system and used during traffic responsive time periods to select the optimal signal timing plan.

• The predicted travel time data can be used as actionable trip planning information for travelers to advise them of anticipated traffic conditions.
Thank You!

Questions?

Shayan Khoshmagham,
Data Scientist, Performance Analytics
sxk@iteris.com
510-292-4148