Probe-Based Volume Estimation Using Machine Learning Techniques

Kaveh Farokhi Sadabadi, Przemyslaw Sekula, Zachary Vander Laan
Center for Advanced Transportation Technology (CATT)
University of Maryland

Presented by:
Zachary Vander Laan

2018 NaTMEC Conference
Irvine, CA
Objectives

• Given the following:
  – Probe volumes (processed from GPS traces of a subset of vehicles),
  – Other archived data (speeds, road geometry, weather, etc.)
  – Continuous count data from select locations

• Can we build a model to accurately estimate statewide volumes?
Volume Estimation: General Approach

Develop and Train Model

- Where? TMC segments associated with continuous count stations
- How? Construct machine learning model to learn relation between input variables and continuous count volumes

Apply model to state road network

- Where? All TMCs on road network
- How? Apply trained model to input variables from any TMC segment on the network

June 13, 2018
Data

Data needed at all TMCs

- GPS probe data (INRIX)
- Probe speeds
- Road characteristics
  - # lanes, speed limit, facility type, etc.
- Weather
- TTI hourly volume estimates (optional)

Data needed only at continuous count stations

- Ground truth count data
  - Used for model training / evaluation
  - Used to estimate probe penetration rate

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Trips</th>
<th>Waypoints</th>
<th>Median Pen. Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maryland</td>
<td>20 M</td>
<td>1.4 B</td>
<td>1.9%</td>
</tr>
<tr>
<td>Florida</td>
<td>75 M</td>
<td>3.4 B</td>
<td>2.1%</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>7 M</td>
<td>595 M</td>
<td>2.3%</td>
</tr>
</tbody>
</table>

Continuous count stations in Florida, with corresponding GPS probe penetration rate

June 13, 2018
Model Evaluation

- **Model**: “Dense” Artificial Neural Network (ANN)
- **Cross validation**: Repeat N times (N = number of stations)
  - Train model using data from all but one continuous count station
  - Generate model predictions using data from remaining station
- **Evaluation**: Compare estimates with actual volumes & generate metrics

June 13, 2018
Quantifying Model Accuracy

\[ y_i = \text{observed volume}, \quad \bar{y}_i = \text{average observed volume}, \quad \hat{y}_i = \text{model volume estimate}, \quad y_{\text{max}} = \text{max observed volume} \]

- **Mean Absolute Percentage Error (MAPE)**
  - Reflects absolute volume accuracy
  - *Good*: 10-15% (high volume), 15-25% (mid volume), 25-??% (low volume)

\[
MAPE = \left( \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right| \right) \times 100
\]

- **Error to Max Flow Ratio (EMFR)**
  - Captures accuracy relative to capacity (max observed flow)
  - < 10% becomes useful, < 5% target

\[
EMFR = \left( \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_{\text{max}}} \right| \right) \times 100
\]

- **Coefficient of Determination (R^2)**
  - Shows explanatory power of model
  - > 0.70 *good*, > 0.80 *better*, > 0.90 *best*

\[
R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}
\]

June 13, 2018
Results: Overview

Summary
- Promising model performance, even across multiple scenarios
- Stable accuracy levels across multiple datasets

Observations
- ↑ Road class = ↑ Accuracy
- ↑ Avg. hourly volume = ↑ Accuracy
- ↑ Avg. hourly GPS counts = ↑ Accuracy

Median Error Metrics by Dataset

<table>
<thead>
<tr>
<th>Dataset</th>
<th>R2</th>
<th>MAPE (%)</th>
<th>EMFR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maryland</td>
<td>0.85</td>
<td>22.6</td>
<td>6.6</td>
</tr>
<tr>
<td>Florida</td>
<td>0.83</td>
<td>24.8</td>
<td>6.6</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>0.82</td>
<td>27.6</td>
<td>7.3</td>
</tr>
</tbody>
</table>
Flagging Unusual Behavior

**Goal:** Develop flags to highlight unusual input data and output model estimates

- **Flag 1** - based on GPS input data (key model “ingredient”)
  - *Typical:* Observed GPS counts within X std. dev of mean GPS counts during same day of week and hour
  - *Low:* Less than *Typical* range
  - *High:* Greater than *Typical* range

- **Flag 2** - based on output model estimates
  - *Typical:* Observed hourly estimates within X std. dev of mean estimates during same day of week and hour
  - *Low:* Less than *Typical* range
  - *High:* Greater than *Typical* range
Statewide Model

- Apply trained model to entire road network in Florida
  - Requires 3 months of hourly input data at ~20k TMCs
  - Generate hourly volume estimates at each input time/location

June 13, 2018
Hourly Freight Volume Estimation

- Use same approach to estimate hourly freight volumes
  - Train model using only ground truth truck counts
  - Need continuous count data by FHWA weight class!

- Initial Florida freight results look promising on high FRC
  - FRC 1 results comparable to all-vehicle model accuracy
  - Not enough data to reliably estimate low FRC

<table>
<thead>
<tr>
<th>FHWA Class 5-13</th>
<th>R²</th>
<th>MAPE (%)</th>
<th>EMFR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.77</td>
<td>37.9</td>
<td>7.5</td>
</tr>
<tr>
<td>FRC 1</td>
<td>0.83</td>
<td>23.5</td>
<td>6.3</td>
</tr>
<tr>
<td>FRC 2</td>
<td>0.76</td>
<td>42.2</td>
<td>7.9</td>
</tr>
<tr>
<td>FRC 3 &amp; 4</td>
<td>0.65</td>
<td>48.9</td>
<td>9.2</td>
</tr>
</tbody>
</table>

Median Error Metrics: Florida Truck Volume Estimation

Similar accuracy to all-vehicle Florida model

June 13, 2018
AADT Estimation

- Possible Approaches:
  - **Method 1**: Aggregate hourly volume estimates
  - **Method 2**: Develop separate AADT estimation model

- Generated initial results via **Method 1**
  - Promising model performance!
  - Consistent with expectations along major highways and urban areas

<table>
<thead>
<tr>
<th>$R^2$</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.86</td>
<td>15</td>
</tr>
</tbody>
</table>

- Future work will compare approaches
Next Steps

**Technical**

- Refine hourly volume estimation models
- Scale approach to statewide networks in MD and NH
- Further investigate hourly truck volume and AADT estimation
- Explore transferability of models between different states

**Program Level**

→ Get out of the lab and operationalize!

June 13, 2018
Questions

Contact Information

Kaveh Farokhi Sadabadi (PI)
kfarokhi@umd.edu

Przemyslaw Sekula
psekula@umd.edu

Zachary Vander Laan
zvanderl@umd.edu

Acknowledgments

• I-95 Corridor Coalition
• MDOT, FDOT, NHDOT

June 13, 2018