Statewide Traffic Volume Estimation using GPS Traces

Presented by:
Zachary Vander Laan
Center for Advanced Transportation Technology (CATT)
University of Maryland

2018 ITS America Annual Meeting
Detroit, MI

June 6, 2018
Presentation Outline

• Overview
  • Objectives
  • Volume estimation approach

• Florida case study
  • Dataset
  • Results
  • Statewide Estimation
  • AADT/AADWT
  • Truck Volumes

• Summary / Next Steps

• Q & A

June 6, 2018
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 6, 2018
Florida Dataset (Q4 2016)

Data needed at all TMCs

- **GPS probe data (INRIX)**
  - 75M trips, 3.4B pts
  - Penetration rate: 2.1% median
  - Snapped to base map

- **Probe speeds (HERE)**

- **Road characteristics**
  - # lanes, speed limit, facility type, etc.

- **Weather**

- **TTI hourly volume estimates**

Data needed only at continuous count stations

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

June 6, 2018
Florida Model Evaluation

- **Model**: “Dense” Artificial Neural Network (ANN)
- **Cross validation** (repeat 173 times)
  - Train model using data from 172 of 173 continuous count stations
  - Generate model predictions using data from remaining station
- **Evaluation**: Compare estimates with actual volumes & generate metrics

June 6, 2018
Quantifying Model Accuracy

\( y_i = \) observed volume, \( \bar{y}_i = \) average observed volume, \( \hat{y}_i = \) model volume estimate, \( y_{\text{max}} = \) max observed volume

• **Mean Absolute Percentage Error (MAPE)**
  - Reflects absolute volume accuracy
  - **Good:** 10-15% (high volume),
    20-25% (mid volume)
    30-50% (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 Capacity (EMFR)**
  - Captures accuracy relative to capacity (max observed flow)
  - \(< 10\% \text{ becomes useful}, \ < 5\% \text{ target}\)

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

• **Coefficient of Determination (R2)**
  - Shows explanatory power of model
  - \(> 0.70 \text{ good}, \ > 0.80 \text{ better}, \ > 0.90 \text{ best}\)

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

June 6, 2018
Florida Results: Summary

**Overall median error metrics:**
- R2 = 0.83
- MAPE = 25%
- EMFR = 7%

**Summary**
Promising model performance, even over a variety of scenarios

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

---

**Median Error Metrics by Scenario**

<table>
<thead>
<tr>
<th>Road Classification</th>
<th>R2</th>
<th>MAPE (%)</th>
<th>EMFR (%)</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRC 1 (Interstates)</td>
<td>0.86</td>
<td>21</td>
<td>6</td>
<td>195704</td>
</tr>
<tr>
<td>Maryland (mostly FRC 1)</td>
<td>0.86</td>
<td>23</td>
<td>7</td>
<td>158040</td>
</tr>
<tr>
<td>FRC 2 (Other Freeways &amp; Expressways)</td>
<td>0.82</td>
<td>26</td>
<td>7</td>
<td>370567</td>
</tr>
<tr>
<td>FRC 3 &amp; 4 (Other principal &amp; minor arterials)</td>
<td>0.83</td>
<td>33</td>
<td>7</td>
<td>128419</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hourly Volume (vph)</th>
<th>R2</th>
<th>MAPE (%)</th>
<th>EMFR (%)</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1k</td>
<td>0.81</td>
<td>29</td>
<td>7</td>
<td>465591</td>
</tr>
<tr>
<td>1k-2k</td>
<td>0.86</td>
<td>22</td>
<td>6</td>
<td>164465</td>
</tr>
<tr>
<td>2k-3k</td>
<td>0.88</td>
<td>18</td>
<td>6</td>
<td>49221</td>
</tr>
<tr>
<td>3k+</td>
<td>0.87</td>
<td>19</td>
<td>6</td>
<td>15413</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Avg probe counts / hr</th>
<th>R2</th>
<th>MAPE (%)</th>
<th>EMFR (%)</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Low&quot; [0-6]</td>
<td>0.78</td>
<td>38</td>
<td>8</td>
<td>214557</td>
</tr>
<tr>
<td>&quot;Medium&quot; [6-17]</td>
<td>0.84</td>
<td>24</td>
<td>7</td>
<td>249730</td>
</tr>
<tr>
<td>&quot;High&quot; [17-145]</td>
<td>0.85</td>
<td>22</td>
<td>6</td>
<td>230403</td>
</tr>
</tbody>
</table>

June 6, 2018
Florida Statewide Model

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

Continuous count station selected that exhibits typical (median) model performance.
AADT & AAWDT Estimation

<table>
<thead>
<tr>
<th>Measure (VPD)</th>
<th>$R^2$</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AADT</td>
<td>0.86</td>
<td>15</td>
</tr>
<tr>
<td>AAWDT</td>
<td>0.87</td>
<td>15</td>
</tr>
</tbody>
</table>

June 6, 2018
Freight Volume Estimation

→ Apply model to estimate hourly freight volumes
  • Leverage highly-granular FDOT continuous count data

→ Initial freight volume results look promising, particularly on higher functional road classes

<table>
<thead>
<tr>
<th>FHWA Class 5-13</th>
<th>R²</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.77</td>
<td>38</td>
</tr>
<tr>
<td>FRC 1</td>
<td>0.83</td>
<td>24</td>
</tr>
<tr>
<td>FRC 2</td>
<td>0.76</td>
<td>42</td>
</tr>
<tr>
<td>FRC 3 &amp; 4</td>
<td>0.65</td>
<td>49</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FHWA Class 7-13</th>
<th>R²</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.66</td>
<td>44</td>
</tr>
<tr>
<td>FRC 1</td>
<td>0.80</td>
<td>26</td>
</tr>
<tr>
<td>FRC 2</td>
<td>0.62</td>
<td>49</td>
</tr>
<tr>
<td>FRC 3 &amp; 4</td>
<td>0.38</td>
<td>54</td>
</tr>
</tbody>
</table>

* Median error metrics
Summary

- Hourly volume estimates:
  - Estimates meet requirements for most planning and operational purposes
  - Estimation quality improves with road class and actual volumes (number of probes)

- AADT and AAWDT estimates:
  - High level of accuracy
  - Consistent with expectations along major highways and urban areas

- Freight volumes
  - Initial results are promising, especially on FRC1 roads. Additional details to come soon.

Next Steps

- Low volume / special event flags
- Explore transferability of models between different states

June 6, 2018
Questions

Contact Information

Zachary Vander Laan
Center for Advanced Transportation Technology (CATT)
University of Maryland, College Park
zvanderl@umd.edu

Acknowledgments

Kaveh Farokhi Sadabadi (PI)
Przemyslaw Sekula
Nikola Markovic

June 6, 2018