Utilizing Real-World Transportation Data for Accurate Traffic Prediction

Pan, Demiryurek, Shahabi, ICDM12

Presented by Liyue Fan
Know the data

- Loop Detectors
- Occupancy, Volume, Speed
- Speed(t)
Goals

- Predict traffic considering rush hours
- Predict traffic when events are present

**Definition 1:** Given a set of observed speed readings $V = \{v_i(j), \ i = 1, \ldots, n; \ j = 1, \ldots, t\}$, where $i$ and $j$ denotes a sensor and continuous time increments, respectively. The prediction problem is to find the set $V = \{v_i(j), \ j = t+1, t+2, \ldots, t+h\}$ for each sensor $i$, where $h$ denotes the prediction horizon. For example, $h=1$ refers to predicting the value of speed at $t+1$, where $t$ represents the current time.
Baseline Methods

- **ARIMA (Auto-Regressive Integrated Moving Average)**

\[ Y_{t+1} = \sum_{i=1}^{p} \alpha_i Y_{t-i+1} + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i+1} + \varepsilon_{t+1} \]

- **HAM (Historical Average Model)**

\[ \nu(t_d, w + h) = \frac{1}{|V(d, w)|} \sum_{s \in V(d, w)} \nu(s) \]
Baseline Limitations

- ARIMA is **not** suitable for long-term prediction
Baseline Limitations (cont.)

- HAM can predict rush hour boundaries
Hybrid Model (H-ARIMA)

- If $\lambda(t) < 0.5$
  - use ARIMA(t)
- Else
  - use HAM(t)

**Algorithm 1** Get $\lambda\{v(j), d, w\}$

**Output:** $\lambda$

1. Let $S = \{V\{v(j), d, w\}\}$
2. Let $Err_{ARIMA} = 0; Err_{HAM} = 0$
3. Initialize ARIMA model with training dataset $\{v(j)\}$
4. $v_{HAM} = \text{Average}(V\{d, w\});$
5. **for all** $v_i \in S$ **do**
   6. $v_{ARIMA} = \text{ARIMA}(i);$  
   7. $Err_{ARIMA} = Err_{ARIMA} + \text{RMSE}(v_i, v_{ARIMA});$
   8. $Err_{HAM} = Err_{HAM} + \text{RMSE}(v_i, v_{HAM});$
5. **end for**
6. $\lambda = \frac{Err_{ARIMA}}{(Err_{ARIMA} + Err_{HAM})}$
7. Return $\lambda.$
Figure 3. Effects of prediction horizons over average $\lambda$
Figure 4. Effects of rush-hour boundaries over $\lambda$
Traffic Event Impact

- Include accidents, constructions, games
- Can cause congestions
Poor Prediction Performance

Figure 5. Impact of an accident on ARIMA and HAM
Historical Event Reports

• Record (Start time, Location, Direction, Type, Affected Lanes)

• Learn Impact Post Mile:

Figure 6. Definition of event impact post-mile
Table I
AVERAGE IMPACT POST-MILE ON EVENT META-ATTRIBUTES

(a) Traffic collision event, affected lanes = 0

<table>
<thead>
<tr>
<th>Location</th>
<th>D</th>
<th>S_{0,4}</th>
<th>S_{4,8}</th>
<th>S_{8,12}</th>
<th>S_{12,16}</th>
<th>S_{16,20}</th>
<th>S_{20,24}</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-405</td>
<td>N</td>
<td>2.07</td>
<td>2.93</td>
<td>3.68</td>
<td>2.92</td>
<td>3.33</td>
<td>1.51</td>
</tr>
<tr>
<td>I-405</td>
<td>S</td>
<td>0.14</td>
<td>3.37</td>
<td>2.61</td>
<td>3.63</td>
<td>4.37</td>
<td>2.03</td>
</tr>
<tr>
<td>I-5</td>
<td>N</td>
<td>0.10</td>
<td>3.32</td>
<td>4.12</td>
<td>4.45</td>
<td>5.51</td>
<td>2.56</td>
</tr>
<tr>
<td>I-5</td>
<td>S</td>
<td>1.17</td>
<td>3.66</td>
<td>3.41</td>
<td>2.43</td>
<td>3.73</td>
<td>1.34</td>
</tr>
</tbody>
</table>
Event Impact Prediction

- For event $e$, sensor $i$:
  - Influence speed decrease $\Delta v_i$ - learned
  - Influence time shift $\Delta t_i$

$$\Delta t_i(e) = \frac{\text{dist}(i, e)}{\text{avg}(\{v_j\})} \text{ where } p(i) \leq p(j) \leq p(e)$$

- Note: HAM is **not** suitable during event time.
H-ARIMA+

• Algorithm:

1) When an event $e$ occurs at time $t$, all the relevant event features (i.e., \{Start-time, Location, Direction, Type, Affected Lanes\}) are incorporated in the EIA model to determine the impact post-mile of $e$.

2) Using the impact post-mile and the location of $e$, the set of all influenced sensors are identified as set \{$s_i$\}.

3) For each sensor $s_i$, during $[t+\Delta t_i(e), t+\Delta t_i(e)+h]$, the predicted value is calculated as $(v_i(t) - \Delta v_i)$, where $h$ is the prediction horizon.

4) After time $t+\Delta t_i(e)+h$, ARIMA is used to predict the rest until the event $e$ is cleared.
Evaluation

(a) Actual speed
(b) MAPE of the road

Figure 11. Case study on I-10 E. segment to Downtown
Evaluation (Cont.)

(a) Actual speed and historical average

(b) MAPE of the sensor

Figure 12. Case study on traffic collision events
Discussion

• Application vs. Research?
• Rough, static estimate for event impact
  – Model traffic flow using multi-sensor reading
  – Update estimate real-time
• Poor overall performance
• Regression of ARIMA/HAM
• After-event Impact
• Event learning