Data Driven Analysis in Transportation Systems

http://imsc.usc.edu/

Ugur Demiryurek, Ph.D.
Associate Director,
Integrated Media Systems Center (IMSC)
Viterbi School of Engineering
University of Southern California
Los Angeles, CA 90089-0781
demiryur@usc.edu
Facilitating an infrastructure for acquiring, processing, storing and querying real-time and historical transportation datasets
Team

• Government
  – LA-Metro: Los Angeles County Metropolitan Transportation Authority
  – RIITS: Regional Integrated Intelligent Transportation System
  – National Science Foundation (NSF)

• USC
  – Price School of Public Policy
  – Stevens Institute
  – Industrial Engineering
  – Marshall School of Business

• Industry
  – Microsoft
  – Intel
  – HP
Rethinking Congestion

Texas Transportation Institute (TTI): Best know source for congestion and mobility studies
Rethinking Congestion

- The reports are static (quarterly)
- Corridors have very different behaviors
- Spatial and temporal granularity of average values are high level (e.g., one hour)
- Not customized for needs of local authorities

### Urban Congestion Report (UCR)

#### A Snapshot of City Congestion Trends for July 2012 through September 2012

<table>
<thead>
<tr>
<th>City</th>
<th>Congested %</th>
<th>Trend Time Index</th>
<th>Planning Time Index</th>
<th>% Change</th>
<th>% Usable Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlanta, GA</td>
<td>n.a.</td>
<td>n.a.</td>
<td>-1.0</td>
<td>-1.0</td>
<td>90%</td>
</tr>
<tr>
<td>Boston, MA</td>
<td>5.59</td>
<td>1.29</td>
<td>1.69</td>
<td>+6%</td>
<td>100%</td>
</tr>
<tr>
<td>Chicago, IL</td>
<td>6.57</td>
<td>1.21</td>
<td>-0.48</td>
<td>-3%</td>
<td>83%</td>
</tr>
<tr>
<td>Detroit, MI</td>
<td>2.82</td>
<td>1.11</td>
<td>-1.72</td>
<td>-5%</td>
<td>100%</td>
</tr>
<tr>
<td>Houston, TX</td>
<td>4.07</td>
<td>2.50</td>
<td>1.50</td>
<td>+2%</td>
<td>92%</td>
</tr>
<tr>
<td>Los Angeles, CA</td>
<td>6.09</td>
<td>1.27</td>
<td>-3.35</td>
<td>-7%</td>
<td>90%</td>
</tr>
<tr>
<td>Minneapolis-St. Paul, MN</td>
<td>4.42</td>
<td>1.23</td>
<td>-2.57</td>
<td>-7%</td>
<td>90%</td>
</tr>
<tr>
<td>Oklahoma City, OK</td>
<td>3.09</td>
<td>1.05</td>
<td>-2.14</td>
<td>-6%</td>
<td>100%</td>
</tr>
<tr>
<td>Orange County, CA</td>
<td>3.51</td>
<td>1.01</td>
<td>-2.50</td>
<td>-7%</td>
<td>90%</td>
</tr>
<tr>
<td>Philadelphia, PA</td>
<td>4.60</td>
<td>1.16</td>
<td>-1.50</td>
<td>-8%</td>
<td>100%</td>
</tr>
<tr>
<td>Phoenix, AZ</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Pittsburgh, PA</td>
<td>5.49</td>
<td>1.22</td>
<td>-3.20</td>
<td>-10%</td>
<td>100%</td>
</tr>
<tr>
<td>Portland, OR</td>
<td>3.53</td>
<td>1.26</td>
<td>-1.67</td>
<td>-8%</td>
<td>90%</td>
</tr>
<tr>
<td>Providence, RI</td>
<td>2.40</td>
<td>1.11</td>
<td>-1.00</td>
<td>-2%</td>
<td>100%</td>
</tr>
<tr>
<td>Riverside - San Bernardino, CA</td>
<td>2.64</td>
<td>1.20</td>
<td>-1.00</td>
<td>-2%</td>
<td>100%</td>
</tr>
<tr>
<td>Sacramento, CA</td>
<td>2.89</td>
<td>1.06</td>
<td>-0.22</td>
<td>-8%</td>
<td>100%</td>
</tr>
<tr>
<td>St. Louis, MO</td>
<td>4.43</td>
<td>1.07</td>
<td>-3.37</td>
<td>-10%</td>
<td>100%</td>
</tr>
<tr>
<td>Salt Lake City, UT</td>
<td>1.46</td>
<td>0.97</td>
<td>-0.49</td>
<td>-6%</td>
<td>60%</td>
</tr>
<tr>
<td>San Antonio, TX</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>San Diego, CA</td>
<td>2.28</td>
<td>1.11</td>
<td>-1.17</td>
<td>-5%</td>
<td>90%</td>
</tr>
<tr>
<td>San Francisco, CA</td>
<td>4.01</td>
<td>1.28</td>
<td>-2.73</td>
<td>-15%</td>
<td>90%</td>
</tr>
<tr>
<td>Seattle, WA</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Tampa, FL</td>
<td>3.13</td>
<td>1.15</td>
<td>-1.99</td>
<td>-4%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Source: U.S. Bureau of Transportation Statistics, Bureau of Transportation Research and Development.
Demo
Rethinking Congestion

• Studied 28 corridors in Los Angeles County
• Evaluated
  – Planning-time index
    • E.g., Planning-time index is 1.60 and the free-flow travel time is 15 minutes, then
      \[ \text{Planning-time} = 15 \times 1.60 = 24 \text{ minutes} \]
  – Travel-time index
  – Hours delayed
  – Gallons wasted
Scatter plots of the TTI measures compared with the IMSC measures. The straight lines indicate where measures by TTI equal measures by IMSC. Total number of data points in each table is 28. The triangle points show the side of the deviation where most observations are.
Rethinking Congestion

• Main findings
  – **TTI Underestimated**: Planning-time index
  – **TTI Overestimated**: Travel-time, delays, gallons wasted

<table>
<thead>
<tr>
<th>Planning Time Index</th>
<th>Travel Time Index</th>
<th>Delay Person-hour</th>
<th>Gallons Wasted</th>
</tr>
</thead>
<tbody>
<tr>
<td>17%</td>
<td>-8%</td>
<td>-23%</td>
<td>-7%</td>
</tr>
</tbody>
</table>

Comparisons of measures between TTI and IMSC dataset
Rethinking Congestion

• There are more extreme congestion conditions than expected during the peak hours
  – Our estimated planning time indexes are 17 percent larger than the TTI estimates
  – Drivers should be prepared to leave more slack time during peak hour trips

• But overall (for 24 hours a day and 7 days a week) congestion is less than expected
  – Annual delay and total congestion cost to be around 2/3 of the stated value from TTI

ClearPath

• The weight of an edge is a function of time, i.e., time-dependent.

• Arrival-time to an edge determines the travel-time on that edge.
ClearPath

• **USC Stevens fund**
  – ClearPath, a spin-off from USC

• **R&D**
  – $2+ million research funding in last 3 years
  – 10 years of research on route planning
  – Access to massive traffic sensor data

• **Intellectual Property**
  – 3 patents: 2 filed and 1 provisional

• **Team**
  – Complementary expertise in engineering and business
## ClearPath

<table>
<thead>
<tr>
<th>Feature</th>
<th>Inrix</th>
<th>Google Maps</th>
<th>NAVTEQ</th>
<th>Waze</th>
<th>ClearPath</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Route guidance</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Shortest distance</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Road-type selector</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Real-time traffic</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>GUI interface</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Static POI</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Crowdsourcing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>Predictive Paths</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√*</td>
</tr>
<tr>
<td>Anticipate event traffic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√*</td>
</tr>
<tr>
<td>FastPath POI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√*</td>
</tr>
</tbody>
</table>

*Patent Pending*
ClearPath

• Value Proposition
  – Improvement over Google Map’s paths: 18%
    • 64 hours per year saved by average commuter
    • $840 Average fuel consumption savings
Demo
ClearPath

- **Goals:**
  - **Exact** fastest path in large time-dependent road networks
  - Fast response time
  - Fast precomputation
  - Low space overhead
  - Efficient update operations


A. Akdogan, U. Demiryurek, C. Shahabi: *Voronoi-Based Geospatial Query Processing with MapReduce*. In IEED CloudCom 2010 (**Best paper award**)

USC Viterbi
School of Engineering
ClearPath

- Dreyfus vs. Time-dependent A*

**Optimality Condition:**

$h(v) = (v, d)$

$h(v)$ should not overestimate the actual distance between $v$ and $d$
Challenge: Finding Heuristic Function $ht(v_i,d) \leq D(v_i,d,t)$ in TD Networks

The distance (travel-time) between any node $v$ and $d$ is time-dependent.
Time-dependent A* Search

- Naïve Heuristic Function:

$$D_{EUC}(v_i, d) \over \text{max(speed)}$$

* Euclidean distance between $v$ and $d$ divided by the maximum speed among the edges

- **Guaranteed** to be a lower-bound as the distance between $v$ and $d$ is never overestimated
- **Problem:** It is a very loose bound, hence yields insignificant performance improvement
Time-dependent A* Search

• Goal:
  – Find a $h(v_i)$ that will never overestimate the time-dependent travel-time between $v_i$ and $d$. This is necessary for Exact results
  – $h(v_i)$ should be as close as possible to actual distances for Efficient processing of fastest path computation
Time-dependent A* Search

• Goal:
  – Find a $h(v_i)$ that will never overestimate the time-dependent travel-time between $v_i$ and $d$. This is necessary for Exact results
  – $h(v_i)$ should be as close as possible to actual distances for Efficient processing of fastest path computation

• Approach:
  – **Step 1:** Partition the road network into non-overlapping partitions (Offline)
  – **Step 2:** Precompute $h(v_i)$ using distances in and between the non-overlapping partitions (Offline)
Time-dependent A* Search

• **Step 1: Partition** the road network using network hierarchies
Time-dependent A* Search

• **Step 2:** Compute *intra and inter distance labels*
  
  – **Intra:** fastest path in *Lower-bound Graph G* from each node $v_i$ to border nodes and border nodes to $v_i$
  
  – **Inter:** fastest path in *Lower-bound Graph G* between border nodes

• Only store the minimum of the intra and inter distance lables

$$LTT(v_i, b_i) = \arg\min(LTT(v_i, b_i), LTT(v_i, b_j))$$

$$LTT(b_i, d) = \arg\min(LTT(b_k, d), LTT(b_i, d))$$

$$LTT(b_i, b_k) = \arg\min(LTT(b_i, b_k), LTT(b_i, b_l), LTT(b_j, b_k), LTT(b_j, b_l))$$
Time-dependent A* Search

• Fast $h(v_i,d)$ computation
  – $h(v_i,d)$ is computed by simple table look-ups (nanoseconds)

![Diagram with nodes and connections]

**Table: Node-to-Partition**

<table>
<thead>
<tr>
<th>Node</th>
<th>Partition</th>
<th>Node-to-Border</th>
<th>Border-to-Node</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_6$</td>
<td>$S_1$</td>
<td>$b_{23},5$</td>
<td>$b_{23},7$</td>
</tr>
<tr>
<td>$n_7$</td>
<td>$S_1$</td>
<td>$b_{4},6$</td>
<td>$b_{9},4$</td>
</tr>
<tr>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
</tr>
<tr>
<td>$n_{16}$</td>
<td>$S_4$</td>
<td>$b_{17},3$</td>
<td>$b_{33},5$</td>
</tr>
<tr>
<td>$n_n$</td>
<td>$S_k$</td>
<td>$b_{u},x$</td>
<td>$b_{v},y$</td>
</tr>
</tbody>
</table>
Future Directions

• More on event impact analysis
  – Games, concerts, Obama visit
• New corridor analysis
  – Light rail opening
  – Express Lane opening
• ClearPath
  – Field tests and surveys
  – New features (Delivery, Multimodal)
Questions?