Data Analytics based Dynamic Passenger Information System

Rakesh Behera
4th year B. Tech. student
Dept. of Civil Engineering
IIT Madras

Devarsh Kumar
4th year dual degree student
Dept. of Civil Engineering
IIT Madras

Lelitha Vanajakshi
Associate Professor
Dept. of Civil Engineering
IIT Madras
Outline

- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Travel time prediction
- Application
- Conclusion
Outline

- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Travel time prediction
- Application
- Conclusion
Public Transportation Issues

Native time taken by passenger = walking time from origin to bus stop + waiting time at the bus stops (original / transfer) + actual travel time on bus + time from bus stop to destination + walking time from destination.
Outline

- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Travel time prediction
- Application
- Conclusion
Proposed Solution

- Bus arrival information system
  - Based on travel time prediction using real-time and historical data
- Trip planner
  - Recommends riders the quickest route along with bus information
Outline

- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Travel time prediction
- Application
- Conclusion
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Input data</th>
<th>Prediction technique</th>
<th>Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bo et al.</td>
<td>One month weekdays bus trip data</td>
<td>Linear regression</td>
<td>Data driven</td>
</tr>
<tr>
<td>Patnaik et al.</td>
<td>6 months weekday data</td>
<td>Regression</td>
<td>Data driven</td>
</tr>
<tr>
<td>Bhandari</td>
<td>7 months AVL data</td>
<td>Time series</td>
<td>Data driven</td>
</tr>
<tr>
<td>Liu et al.</td>
<td>82 days volume and flow data</td>
<td>ANN</td>
<td>Data driven</td>
</tr>
<tr>
<td>Krishnan et al.</td>
<td>15 minute aggregate flow data for 3 months</td>
<td>k – Nearest Neighbor</td>
<td>Model based</td>
</tr>
<tr>
<td>Vanajakshi et al.</td>
<td>Preceding 2 buses data</td>
<td>Kalman filtering</td>
<td>Model based</td>
</tr>
<tr>
<td>Dailey et al.</td>
<td>Different days trips at the same time of the day</td>
<td>Kalman filtering</td>
<td>Model based</td>
</tr>
</tbody>
</table>
Outline

- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Travel time prediction
- Application
- Conclusion
85 days GPS data for 2 buses chosen for demo.

Buses belonged to two routes sharing common end points.

Sampling rate: 6 logs/minute.

Each day had approx. 8000 GPS logs.

Each record had a timestamp, latitude and longitude of the location of the bus.
## Sample Raw Data

<table>
<thead>
<tr>
<th>Timestamp (dd-mm-yyyy HH:MM:SS)</th>
<th>Longitude (degrees)</th>
<th>Latitude (degrees)</th>
</tr>
</thead>
<tbody>
<tr>
<td>17-09-2013 09:30:22</td>
<td>80.12769318</td>
<td>12.92300034</td>
</tr>
<tr>
<td>17-09-2013 09:30:32</td>
<td>80.12789917</td>
<td>12.92298985</td>
</tr>
<tr>
<td>17-09-2013 09:30:43</td>
<td>80.12844849</td>
<td>12.92292023</td>
</tr>
<tr>
<td>17-09-2013 09:30:53</td>
<td>80.12899781</td>
<td>12.92286968</td>
</tr>
<tr>
<td>17-09-2013 09:31:03</td>
<td>80.12966156</td>
<td>12.92282963</td>
</tr>
</tbody>
</table>
Routes Chosen

Figure 1. GPS data of two routes plotted on Google Earth
Data Preprocessing

- The data were transformed into traffic data e.g. travel time, distance, speed.
- Distances were calculated using haversine formula.
- The data were split into sets belonging to individual trips.
- Historical travel times on each 500 m segment of each route were calculated.
## Processed Data

<table>
<thead>
<tr>
<th>Timestamp (dd-mm-yyyy HH:MM:SS)</th>
<th>Longitude (degrees)</th>
<th>Latitude (degrees)</th>
<th>TimeDiff (s)</th>
<th>Dist (m)</th>
<th>CumulDist (m)</th>
<th>Speed (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-09-2013 09:30:22</td>
<td>80.12769318</td>
<td>12.92300034</td>
<td>10</td>
<td>43.11</td>
<td>294.16</td>
<td>4.31</td>
</tr>
<tr>
<td>-09-2013 09:30:32</td>
<td>80.12789917</td>
<td>12.92298985</td>
<td>10</td>
<td>22.38</td>
<td>316.54</td>
<td>2.24</td>
</tr>
<tr>
<td>-09-2013 09:30:43</td>
<td>80.12844849</td>
<td>12.92292023</td>
<td>11</td>
<td>60.11</td>
<td>376.65</td>
<td>5.46</td>
</tr>
<tr>
<td>-09-2013 09:30:53</td>
<td>80.12899781</td>
<td>12.92286968</td>
<td>10</td>
<td>59.87</td>
<td>436.52</td>
<td>5.99</td>
</tr>
<tr>
<td>-09-2013 09:31:03</td>
<td>80.12966156</td>
<td>12.92282963</td>
<td>10</td>
<td>72.17</td>
<td>508.69</td>
<td>7.22</td>
</tr>
</tbody>
</table>
Travel Time Database

Route1

Route2

Route3

tripID, seg1, seg2, seg3
trip1, 85.53, 84.26, 86.45
trip2, 87.78, 85.98, 87.14
trip3, 88.23, 86.29, 87.87
Outline

- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Travel time prediction
- Application
- Conclusion
Clustering

Historical travel times on segments were clustered using the V-clustering algorithm (Lee et al. 2012).

**Brief explanation of V – clustering:**

- For a segment, historical travel times are sorted.
- The list is split at an element that minimizes weighted average variance.
- The child lists are split recursively until a minimum of 25 travel times in each list.

\[
WAV^i = \frac{L_1^i}{L} \cdot \text{Var}(L_1^i) + \frac{L_2^i}{L} \cdot \text{Var}(L_2^i)
\]

This led to the reduction of search space for prediction algorithm.
Historical Clusters Database

Route1
Route2
Route3

Seg1
Seg2
Seg3

Clust1
Clust2

88.24, trip1
87.76, trip15
87.88, trip34

Average Travel Time = 87.5
Travel time prediction

Prediction of travel time on upcoming segments:

1. Past segment travel times are calculated.
2. The cluster to which these travel times should belong to is searched for.
3. The cluster found suggests historical trajectories similar to the current one.
4. Travel times on upcoming segments are calculated by averaging corresponding values from similar historical trajectories.
Outline

- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Data analysis
- Travel time prediction
- Application
- Conclusion
Application

- Bus arrival information system
  - Predicted travel times on the segments are added up to obtain arrival times at bus stops.
  - Predictions update every one minute as the bus moves from one segment to another.

- Trip planner
  - The rider inputs the origin and destination.
  - The algorithm suggests the quickest route by finding the minimum weight (time) path from origin to destination.
Outline

- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Data analysis
- Travel time prediction
- Application
- Conclusion
Conclusion

- Classification of historical trips and clustering of travel times led to faster real-time predictions.
- Use of optimal number of previous segments led to more accurate predictions.
- Smaller interval between updates provides the rider with the latest information.
- The promising results indicate that the approach can be implemented under Indian traffic scenarios.
Contact

- Rakesh Behera
  - Email: rakesh8248@gmail.com
  - Phone: +919884995394

- Devarsh Kumar
  - Email: devarsh10iitm@gmail.com
  - Phone: +919884992880

- Lelitha Vanajakshi
  - Email: lelitha@iitm.ac.in
Thank you
Questions
Outline

- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Data analysis
- Travel time prediction
- Results
- Application
- Conclusion
Results

- Measures of accuracy
  - MAPE, MPE, RMSE
- Histograms of prediction errors were plotted.
- Effect of considering varying number of previous segments was analyzed using error distributions.
<table>
<thead>
<tr>
<th>tripID</th>
<th>seg12</th>
<th>seg23</th>
<th>seg34</th>
</tr>
</thead>
<tbody>
<tr>
<td>trip1</td>
<td>170.53</td>
<td>172.26</td>
<td>174.45</td>
</tr>
<tr>
<td>trip2</td>
<td>175.78</td>
<td>173.98</td>
<td>177.14</td>
</tr>
<tr>
<td>trip3</td>
<td>178.23</td>
<td>177.29</td>
<td>176.87</td>
</tr>
</tbody>
</table>
Error distributions

Figure 5. Absolute percentage errors by considering 1 previous segment.

Figure 6. Absolute percentage errors by considering 2 previous segments.

Figure 7. Absolute percentage errors by considering 3 previous segments.
### Results

<table>
<thead>
<tr>
<th>MSL</th>
<th>MPE (%)</th>
<th>MAPE (%)</th>
<th>RMSE (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.80</td>
<td>12.02</td>
<td>9.22</td>
</tr>
<tr>
<td>2</td>
<td>0.75</td>
<td>8.72</td>
<td>7.20</td>
</tr>
<tr>
<td>3</td>
<td>1.78</td>
<td>9.19</td>
<td>8.08</td>
</tr>
<tr>
<td>4</td>
<td>1.89</td>
<td>9.38</td>
<td>8.26</td>
</tr>
<tr>
<td>5</td>
<td>1.96</td>
<td>9.76</td>
<td>8.46</td>
</tr>
<tr>
<td>6</td>
<td>1.93</td>
<td>10.10</td>
<td>8.68</td>
</tr>
<tr>
<td>7</td>
<td>1.99</td>
<td>10.05</td>
<td>8.40</td>
</tr>
</tbody>
</table>
Measures of Accuracy

- **MAPE**: Mean Absolute Percentage Error

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{T}_i - T_i}{T_i} \right| \times 100
\]

- **MPE**: Mean Percentage Error

\[
MPE = \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{T}_i - T_i}{T_i} \times 100
\]

- **RMSE**: Root Mean Square Error

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( \frac{\hat{T}_i - T_i}{T_i} \right)^2}
\]
Outline

- Public transportation issues
- Proposed solution
- Literature review
- Data and pre-processing
- Data analysis
- Travel time prediction
- Results
- Application
- Conclusion
Data Analysis

- Used for classifying historical trips into groups of similar characteristics.

Analyses

- Within – day travel time variation
- Weekdays versus weekends
- Monthly variation
Within – day variation

Figure 2. Space-time trajectories for south bound trips on route B51 in Chennai showing within – day variation.
Weekdays versus weekends

Figure 3. Space-time trajectories for south bound trips on route V51 in Chennai comparing weekend and weekday trips.
Figure 4. Space-time trajectories for south bound trips on route B51 in Chennai showing monthly variation.
Haversine Formula

\[ D = R \ast \cos^{-1}(a + b) \]

where

\[ D = \text{distance between two GPS points} \]
\[ R = \text{mean radius of Earth} = 6371000 \text{ m} \]

\[ a = \cos\left(\frac{\pi}{2} - \text{lat}_1\right) \cos\left(\frac{\pi}{2} - \text{lat}_2\right) \]

\[ b = \sin\left(\frac{\pi}{2} - \text{lat}_1\right) \sin\left(\frac{\pi}{2} - \text{lat}_2\right) \cos(\text{long}_1 - \text{long}_2) \]

\[ \text{lat}_1 = \text{latitude of the first GPS point} \]
\[ \text{lon}_1 = \text{longitude of the first GPS point} \]
\[ \text{lat}_2 = \text{latitude of the second GPS point} \]
\[ \text{lon}_2 = \text{longitude of the second GPS point} \]