Bus Travel Time Predictions using Additive Models

Matthías Kormáksson
IBM Research - Brazil

December 4, 2014
Overview

- Problem Statement
- Motivating Data
- Background
- Proposed Solution
- Experiments
- Conclusions
Given current time and location of the bus, what are the arrival times at subsequent bus stops?
Motivating Data

- **GPS** (lat, lon, timestamp) collected from public buses in the city of Rio de Janeiro.
- **Bus Route Data** provide piecewise linear representations of routes and bus stop locations.
- **Normalization**
  Map GPS coordinates onto a 1-dimensional scale measuring distance from origin.
- **Travel Times**
  inferred by differences between consecutive time stamps.
Cumulative Space-Time Trajectories

- GPS mapped onto a cumulative distance scale (x-axis).
- Cumulative travel time (y-axis) defined as zero at origin.
- Interpolation is required to infer time “zero”.
- Scatter plots show raw measurements observed at irregular spatial locations.
Let $0 = p_0 < p_1 < \cdots < p_K$ denote the bus stops.

- Cumulative space-time trajectories normalized at 0 consist of cumulative distances $0 \leq dist_{ij} \leq p_K$, and cumulative travel times $T_{ij}$ (bus $i$, obs $j$).

- To make future predictions for bus located at $p_k$ we normalize trajectories at $p_k$ (with $dist=0$ and $T=0$ at $p_k$).
Consider the simple scatterplot smoothing model

\[ y_i = f(x_i) + \varepsilon_i. \]

Represent the function as

\[ f(x) = \sum_{j=1}^{q} \beta_j \phi_j(x), \]

where \( \phi_j(x) \) are known basis functions and \( \beta_j \) are coefficients to be estimated. An intuitive example is the piecewise linear representation, involving basis functions

\[ \phi_1(x) = 1, \quad \phi_2(x) = x, \quad \phi_{j+2}(x) = (x - \tau_j)_+, \]

where \((x - \tau_j)_+ = \max(0, x - \tau_j)\), and \(\tau_j\) are called knots.
Piecewise linear spline functions

By setting up the design matrix appropriately the spline smoothing becomes a linear model:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 & (x_1 - \tau_1)_+ & \cdots & (x_1 - \tau_K)_+ \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & (x_n - \tau_1)_+ & \cdots & (x_n - \tau_K)_+ \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_q \end{pmatrix}. $$

Estimated by minimizing the least squares:

$$\min_{\beta} \left\| y - X\beta \right\|^2.$$
Cubic regression spline functions

By replacing the piecewise linear term with a cubic term in the design matrix, the model becomes:

\[
\begin{pmatrix}
y_1 \\
\vdots \\
y_n
\end{pmatrix}
= 
\begin{pmatrix}
1 & x_1 & |x_1 - \tau_1|^3 & \cdots & |x_1 - \tau_K|^3 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_n & |x_n - \tau_1|^3 & \cdots & |x_n - \tau_K|^3
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\vdots \\
\beta_q
\end{pmatrix}.
\]

Estimated by minimizing the least squares:

\[
\min_{\beta} \|y - X\beta\|^2.
\]
Penalized spline smoothing

By introducing a smoothness parameter $\lambda$ we can control the smoothness of the model

$$
\begin{pmatrix}
  y_1 \\
  \vdots \\
  y_n
\end{pmatrix} = \begin{bmatrix}
  1 & x_1 & |x_1 - \tau_1|^3 & \cdots & |x_1 - \tau_K|^3 \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  1 & x_n & |x_n - \tau_1|^3 & \cdots & |x_n - \tau_K|^3
\end{bmatrix} \begin{pmatrix}
  \beta_1 \\
  \vdots \\
  \beta_q
\end{pmatrix}.
$$

We minimize the penalized least squares and estimate $\lambda$ by Generalized Cross Validation (GCV):

$$
\min_{\beta} \|y - X\beta\|^2 + \lambda \beta' D \beta, \quad D = diag(0, 0, 1, \ldots, 1)
$$
Penalized tensor product smoothing

Consider the bivariate smoothing model

\[ y_i = f(x_{1i}, x_{2i}) + \varepsilon_i. \]

Represent the function as

\[ f(x_1, x_2) = \sum_{j=1}^{q_1} \sum_{k=1}^{q_2} \beta_{jk} \phi_j(x_1) \psi_k(x_2), \]

where \( \phi_j(x) \) and \( \psi_k(x) \) are known basis functions (e.g. linear/cubic splines) and \( \beta_{jk} \) are coefficients to be estimated.

By appropriately specifying the design matrix in terms of the basis functions the above model becomes linear:

\[ y = X\beta + \varepsilon. \]
Proposed solution - Additive Models

Model 1: Basic Additive Model (BAM)

\[ T_{ij} = \beta_0 + f_1(\text{dist}_{ij}) + f_2(\text{time}_i) + f_3(\text{dist}_{ij}, \text{time}_i) + \epsilon_{ij}. \]

**Estimation:** Represent \( f_1, f_2 \) with penalized splines and \( f_3 \) with penalized tensor product. Model becomes linear in parameters.

- ... allow the linear predictor to depend on unknown smooth functions of predictor variables.
- (Left): Trajectories of route 121 stratified by hour.
- Note how travel time changes smoothly as function of distance.
- Note how functional relationship changes with the hour, \( \text{time}_i \).
Model 1: Basic Additive Model (BAM)

\[ T_{ij} = \beta_0 + f_1(dist_{ij}) + f_2(time_i) + f_3(dist_{ij}, time_i) + \varepsilon_{ij}, \]

Summary: All terms are significant and adjusted \( R^2 = 0.903 \).
Extended Additive Model

Additive Models are flexible and allow for additional linear predictors.

### Model 2: Extended Additive Model (EAM)

\[
T_{ij} = \beta_0 + \beta_1 \cdot \text{weekend}_i + f_1(\text{dist}_{ij}, \text{weekend}_i) + \beta_2 \cdot T_{ij}^{\text{last}} + f_2(\text{time}_i) + f_3(\text{dist}_{ij}, \text{time}_i) + \varepsilon_{ij}.
\]

**Summary:** All terms are significant and adjusted \( R^2 = 0.919 \).
Recall that time zero for bus \( i \) was inferred by interpolating two consecutive time stamps before and after origin.

**Model 3: Additive Mixed Model (AMM)**

\[
T_{ij} = \beta_0 + b_{0i} + \beta_1 \cdot \text{weekend}_i + f_1(\text{dist}_{ij}, \text{weekend}_i) + \\
+ \beta_2 \cdot T_{ij}^{\text{last}} + f_2(\text{time}_i) + f_3(\text{dist}_{ij}, \text{time}_i) + \epsilon_{ij},
\]

where \( b_{0i} \sim N(0, \sigma_b^2) \) is a corrective random intercept.

**Summary:** \( \hat{\sigma}_b = 3 \) minutes and adjusted \( R^2 = 0.968 \).
Experiments

Data: Four bus routes: 603, 627, 862, and 121.

Test data: 14 random days (out of 3 months of data).

Training data: 10, 20, 30 days before each test date.

Prediction: made for all buses \(i\) in test set from every bus stop \(p_k\) until end of route.

Error: calculated for bus \(i\) at observed \(dist_{ij}\) and stratified by prediction distance \(|dist_{ij} - p_k|\).

<table>
<thead>
<tr>
<th>Route</th>
<th># trajectories</th>
<th># stops</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>603</td>
<td>1,276</td>
<td>15</td>
<td>4km</td>
</tr>
<tr>
<td>627</td>
<td>1,325</td>
<td>54</td>
<td>15km</td>
</tr>
<tr>
<td>862</td>
<td>7,882</td>
<td>24</td>
<td>10km</td>
</tr>
<tr>
<td>121</td>
<td>2,515</td>
<td>18</td>
<td>15km</td>
</tr>
</tbody>
</table>
Experimental results

Box plots of absolute prediction errors (min)

Method
- Kernel
- BAM
- EAM
- AMM

Cumulative distance from origin (km)

Route 603
Route 627
Route 862
Route 121

© 2014 IBM Corporation
Experimental results

Table: Mean Absolute Relative Error (MARE)

<table>
<thead>
<tr>
<th>Route</th>
<th># days</th>
<th>Method</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>BAM</td>
<td>EAM</td>
<td>AMM</td>
<td>Kernel</td>
<td>SVM</td>
</tr>
<tr>
<td>603</td>
<td>10</td>
<td>19.9%</td>
<td>19.7%</td>
<td>18.4%</td>
<td>21.3%</td>
<td>64.4%</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>20.1%</td>
<td>19.8%</td>
<td>18.5%</td>
<td>21.3%</td>
<td>64.7%</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>19.8%</td>
<td>19.6%</td>
<td>18.3%</td>
<td>21.3%</td>
<td>64.8%</td>
</tr>
<tr>
<td>627</td>
<td>10</td>
<td>16.3%</td>
<td>14.7%</td>
<td>13.8%</td>
<td>18.1%</td>
<td>28.8%</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>15.2%</td>
<td>14.2%</td>
<td>13.4%</td>
<td>17.3%</td>
<td>30.0%</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>15.1%</td>
<td>14.0%</td>
<td>13.2%</td>
<td>17.1%</td>
<td>29.4%</td>
</tr>
<tr>
<td>862</td>
<td>10</td>
<td>22.1%</td>
<td>19.5%</td>
<td>18.0%</td>
<td>23.8%</td>
<td>26.4%</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22.5%</td>
<td>19.3%</td>
<td>18.0%</td>
<td>23.6%</td>
<td>26.8%</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>22.2%</td>
<td>19.3%</td>
<td>17.9%</td>
<td>23.4%</td>
<td>25.6%</td>
</tr>
<tr>
<td>121</td>
<td>10</td>
<td>23.1%</td>
<td>20.9%</td>
<td>19.2%</td>
<td>23.9%</td>
<td>41.5%</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22.9%</td>
<td>20.7%</td>
<td>19.1%</td>
<td>23.6%</td>
<td>41.4%</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>22.7%</td>
<td>20.3%</td>
<td>18.9%</td>
<td>23.4%</td>
<td>41.2%</td>
</tr>
</tbody>
</table>

\[
MARE = \frac{1}{N} \sum_{i,j} \left| T_{ij} - \hat{T}_{ij} \right| / T_{ij}
\]
Conclusions

- Proposed solution:
  - models travel times directly using raw irregular GPS data.
  - models spatial and temporal effects through smooth functions thus avoiding any discretization.
  - allows for flexible incorporation of additional predictors.

- We showed that by including a random intercept we correct for an interpolation error.

- Demonstrated on a large real-world GPS data that our method achieved superior performance (as compared to existing methods).
Other Projects at IBM - Oil&Gas

Field with well logs from vertical wells and production data from horizontal wells

Well log data

Model Building

Functional Principal Component Analysis

Kriging

Multiple Linear Regression

Prediction (color maps)
Other Projects at IBM - Oil&Gas


Contact: matkorm@br.ibm.com