Modeling Route Choice Behavior From Smartphone GPS data

Michel Bierlaire

TRANS-P-OR, EPFL

August 19, 2010
Objective: estimate route choice models from GPS data
method: GPS data $\Rightarrow$ path observations $\Rightarrow$ route choice behavior
challenge:
- sparsity and inaccuracy of GPS data
- ambiguity with respect to the real path
Sparsity and inaccuracy of the smartphone GPS data

Blue: GPS device  Red: Nokia N95
Sparsity and inaccuracy of the smartphone GPS data

Blue: GPS device  Red: Nokia N95
Deterministic map matching introduces biases
Network-free data

Bierlaire and Frejinger (2008)

Contribution to the likelihood:

\[
\Pr(\hat{x}_1, \ldots, \hat{x}_T | S) = \sum_{s \in S} \Pr(s | S) \sum_{p \in C(s)} \Pr(\hat{x}_1, \ldots, \hat{x}_T | p) \Pr(p | C(s); \beta),
\]
Measurement model $\Pr(\hat{x}_1, \ldots, \hat{x}_T | p)$

Probability that a vehicle traveling along $p$ generates the GPS trace

Decomposition:

$$\Pr(\hat{x}_1, \ldots, \hat{x}_T | p) = \Pr(\hat{x}_T | \hat{x}_1, \ldots, \hat{x}_{T-1}, p) \Pr(\hat{x}_1, \ldots, \hat{x}_{T-1} | p).$$

The recursion starts with:

$$\Pr(\hat{x}_1 | p) = \int_{x_1 \in p} \Pr(\hat{x}_1 | x_1, p) \Pr(x_1 | p) dx_1,$$

where

- $\Pr(x_1 | p) = 1/L_p$: prior on the location
- $\Pr(\hat{x}_1 | x_1, p) = \Pr(\hat{x}_1 | x_1)$: measurement error of the device.
Measurement error of the device: \( \Pr(\hat{x}|x) \)

- Assumption: the latitudinal and longitudinal errors are i.i.d. normal with variance \( \sigma^2 \).
- Therefore, the measurement error follows a Rayleigh distribution

\[
\Pr(\hat{x}_1|x_1) = \Pr(\text{error} \leq \|\hat{x}_1 - x_1\|_2) = \exp \left( -\frac{\|\hat{x}_1 - x_1\|_2^2}{2\sigma^2} \right).
\]

- \( \sigma \) accounts also for errors in network coding
Iteration $k$

\[
\Pr(\hat{x}_k|\hat{x}_1, \ldots, \hat{x}_{k-1}, p) = \\
\int_{x_k \in \mathcal{P}} \Pr(\hat{x}_k|x_k, \hat{x}_1, \ldots, \hat{x}_{k-1}, p) \Pr(x_k|\hat{x}_1, \ldots, \hat{x}_{k-1}, p) \, dx_k.
\]

- $\Pr(\hat{x}_k|x_k, \hat{x}_1, \ldots, \hat{x}_{k-1}, p) = \Pr(\hat{x}_k|x_k)$: measurement error.
- $\Pr(x_k|\hat{x}_1, \ldots, \hat{x}_{k-1}, p) = \Pr(x_k|\hat{x}_{k-1}, p)$ predicts the position at time $\hat{t}_k$ of the traveler.
Position of the traveler

\[ \Pr(x_k|\hat{x}_{k-1}, p) \]

\[ = \int_{x_{k-1} \in p} \Pr(x_k|x_{k-1}, \hat{x}_{k-1}, p) \Pr(x_{k-1}|\hat{x}_{k-1}, p) dx_{k-1}. \]

\[ = \int_{x_{k-1} \in p} \Pr(x_k|x_{k-1}, \hat{t}_{k-1}, \hat{t}_k, p) \frac{\Pr(\hat{x}_{k-1}|x_{k-1}, p)}{\int_{x_{k-1} \in p} \Pr(\hat{x}_{k-1}|x_{k-1}, p)} dx_{k-1}. \]

- \( \Pr(\hat{x}_{k-1}|x_{k-1}, p) \): measurement error
- \( \Pr(x_k|x_{k-1}, \hat{t}_{k-1}, \hat{t}_k, p) \): movement model
Movement model $\text{Pr}(x|x^-, t^-, t, p)$

$x$ is the position of the device at time $t$ if the position at time $t^-$ is $x^-$, and the device is traveling along path $p$.

Random variable with pdf:

$$f_x(x|x^-, t^-, t, p) = f_v \left( \frac{d_p(x^-, x)}{t - t^-} \right)$$

- $v$: traveling speed of the device.
- proposed distribution:
  - $w \times \text{Negative Exponential} + (1 - w) \times \text{Lognormal}$
- Mixture of two regimes: stop and go
- distribution of $v$ estimated from observed GPS data.
Model fitting
Model estimates

\[ f_v (v) = w \lambda \exp^{-\lambda v} + (1 - w) \frac{1}{\sqrt{\pi \sigma^2}} \exp\left(-\frac{(\ln v - \mu)^2}{2\sigma^2}\right), \]

Speed records: 658.

<table>
<thead>
<tr>
<th>parameter</th>
<th>estimate</th>
<th>standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>( w )</td>
<td>0.528</td>
<td>0.0362</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>0.041</td>
<td>0.0032</td>
</tr>
<tr>
<td>( \mu )</td>
<td>3.843</td>
<td>0.0206</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.250</td>
<td>0.0200</td>
</tr>
</tbody>
</table>

Parameters estimated by R.
Summary

At this point,

$$Pr (\hat{x}_1, \ldots, \hat{x}_T | p)$$

can be computed combining

- a device measurement error model
- a movement model

But it involves complex integrals.
Computing integrals $I = \int_{x \in p} f(x) \, dx$

Truncate the domain of the integrals in each GPS point’s DDR:

$$\exp \left( -\frac{||\hat{x} - x||^2}{2\hat{\sigma}^2} \right) \geq \theta$$

$\theta = 0.65$ and $\hat{\sigma} = 104.4 = \sqrt{100^2 + 30^2}$
Illustration: the actual path (0.08)
Illustration: another path (0.04)
Illustration: map matching result (0)
Path generation algorithm

- Find DDR $D_1$ and associated link set $L_1$ for the first GPS point.
- generate $P_1$ from $L_1$.
- At each iteration $k$:
  - Build bounded shortest path trees from end nodes of $P_{k-1}$
  - Generate $L_k$.
  - Combine $P_{k-1}$, $L_k$ and the shortest path trees to form $P_k$
  - Eliminate unlikely paths from $P_k$
Illustration: path generation algorithm
## Details of a trip

<table>
<thead>
<tr>
<th>path id</th>
<th>origin id</th>
<th>destination id</th>
<th>likelihood</th>
<th>normalized likelihood</th>
<th>length (km)</th>
<th>path size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>289488138</td>
<td>287096130</td>
<td>0.01</td>
<td>0.02</td>
<td>1.65</td>
<td>0.03</td>
</tr>
<tr>
<td>2</td>
<td>289488138</td>
<td>287095960</td>
<td>0.01</td>
<td>0.01</td>
<td>1.62</td>
<td>0.03</td>
</tr>
<tr>
<td>3</td>
<td>313144339</td>
<td>287095960</td>
<td>0.07</td>
<td>0.12</td>
<td>1.66</td>
<td>0.04</td>
</tr>
<tr>
<td>4</td>
<td>313144339</td>
<td>287095960</td>
<td>0.06</td>
<td>0.10</td>
<td>1.70</td>
<td>0.04</td>
</tr>
<tr>
<td>5</td>
<td>313144339</td>
<td>287096130</td>
<td>0.09</td>
<td>0.14</td>
<td>1.74</td>
<td>0.04</td>
</tr>
<tr>
<td>6</td>
<td>313144339</td>
<td>287096130</td>
<td>0.07</td>
<td>0.12</td>
<td>1.78</td>
<td>0.05</td>
</tr>
<tr>
<td>7</td>
<td>313144339</td>
<td>287095960</td>
<td>0.06</td>
<td>0.10</td>
<td>1.70</td>
<td>0.05</td>
</tr>
<tr>
<td>8</td>
<td>313144339</td>
<td>287096130</td>
<td>0.10</td>
<td>0.16</td>
<td>1.69</td>
<td>0.03</td>
</tr>
<tr>
<td>9</td>
<td>313144339</td>
<td>287096130</td>
<td>0.08</td>
<td>0.14</td>
<td>1.73</td>
<td>0.03</td>
</tr>
<tr>
<td>10</td>
<td>313144339</td>
<td>287095960</td>
<td>0.05</td>
<td>0.09</td>
<td>1.74</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Number of GPS points: 16, travel time: 3min
Route choice model with probabilistic observations

\[ \Pr(\hat{x}_1, \ldots, \hat{x}_T | S) = \sum_{s \in S} \Pr(s | S) \sum_{p \in C(s)} \Pr(\hat{x}_1, \ldots, \hat{x}_T | p) \Pr(p | C(s); \beta), \]

- \( \Pr(s | S) = 1/\#S \);
- \( \Pr(\hat{x}_1, \ldots, \hat{x}_T | p) \): measurement model
- \( \Pr(p | C(s); \beta) \): route choice model.
Model specification

• Path Size Logit (PSL):

\[ V_p = \beta_{PS} \ln PS_p + \beta_{\ell} \ln \text{Length}_p \]  

(1)

\[ PS_p = \sum_{a \in p} \frac{l_a}{l_p} \frac{1}{\sum_{q \in C(s)} \delta_{aq}} \]  

(2)

• Alternative sampling: biased random walk: 50 draws, Kumaraswamy parameters \( b_1 = 30 \) and \( b_2 = 0.4 \)

• Data: 17 car trips of one smartphone user
Model estimates

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Value</th>
<th>Rob. Std. Error</th>
<th>Rob. t-test</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{PS}$</td>
<td>3.95</td>
<td>1.40</td>
<td>2.81</td>
<td>0.00</td>
</tr>
<tr>
<td>$\beta_{I}$</td>
<td>-76.2</td>
<td>36.8</td>
<td>-2.07</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Number of observations: 17
Null log-likelihood: -66.008
Final log-likelihood: -33.975
Adjusted rho-square: 0.455
Model estimated by BIOGEME
Conclusions

- The measurements likelihood model is meaningful.
- The path generation algorithm is suitable for smartphone data.
- The estimated route choice behavior is reasonable.