Route choice with smartphones GPS data

Michel Bierlaire

transp-or.epfl.ch

Transport and Mobility Laboratory, EPFL
Motivation

- Travelers are equipped with smartphones
- Smartphones are equipped with sensors
- Can we learn mobility patterns from them?
Objectives

- Focus on GPS data from smartphone
- Reconstruct actual paths
- Model route choice behavior
Issues
Issues
Issues

- Low data collection rate to save battery
- Inaccuracy due to technological constraints
- Smartphone carried in bags, pockets: weaker signal
- Map matching algorithms do not work with this data
Context

- Network: $G = (N, A)$
- Node coordinates: $x_n = \{\text{lat, lon}\}$
- Arc geometry:
  \[
  \mathcal{L}_a : [0, 1] \rightarrow \mathbb{R}^2.
  \]
  Example: straight line
  \[
  \mathcal{L}_a (\ell) = (1 - \ell) x_u + \ell x_d.
  \]
- Model for the movement of the mobile phone:
  \[
  x = S(x^-, t^-, t, p)
  \]
  - Ideally a traffic simulator
  - Simpler models are used in practice
  - Random variable with density $f_x(x|x^-, t^-, t, p)$
Data

One measurement: \( \hat{g} = \left( \hat{t}, \hat{x}, \hat{\sigma}^x, \hat{v}, \hat{\sigma}^v, \hat{h} \right) \),

- \( \hat{t} \), a time stamp;
- \( \hat{x} = (\hat{x}_{\text{lat}}, \hat{x}_{\text{lon}}) \), a pair of coordinates;
- \( \hat{\sigma}^x \), the standard deviation of the horizontal error in the location measurement;
- \( \hat{v} \), a speed measurement (km/h) and,
- \( \hat{\sigma}^v \), the standard deviation of the error in that measurement;
- \( \hat{h} \), a heading measurement, that is the angle to the north direction, from 0 to 359, clockwise.

Sequence: \( (\hat{g}_1, \ldots, \hat{g}_T) \)
Measurement equations

Objective:

- Given a path $p$
- Given a sequence $(\hat{g}_1, \ldots, \hat{g}_T)$
- What is the likelihood that the sequence has been generated by a smartphone moving along path $p$?
- Note: different approach from map matching, which is essentially a projection procedure.
- We focus on the position only
- We derive

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

- ... recursively

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

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

- integral spans all locations \( x_1 \) on path \( p \)
- no prior information on \( x_1 \)

\[
Pr(x_1|p) = 1/L_p
\]

- a smarter way would be to assign more probability in the beginning of the path
- measurement error of the device:

\[
Pr(\hat{x}_1|x_1, p) = Pr(\hat{x}_1|x_1)
\]
Measurement error of the device

- Assume that latitudinal and longitudinal errors are i.i.d. normal with variance $\sigma^2$
- Measurement error is Rayleigh
- $\sigma^2$ unknown, estimate:

$$\hat{\sigma}^2 = \sigma_{\text{network}}^2 + (\hat{\sigma}_1^x)^2$$

where
- $\sigma_{\text{network}}^2$: network coding errors
- $(\hat{\sigma}_1^x)^2$: GPS errors.

$$\Pr(\hat{x}_1 | x_1) = \exp \left(-\frac{\|\hat{x}_1 - x_1\|_2^2}{2\hat{\sigma}^2}\right).$$
Recursion: first step

\[ \Pr(\hat{x}_1|p) = \frac{1}{L_p} \int_{x_1} \exp \left( - \frac{\|\hat{x}_1 - x_1\|^2}{2\hat{\sigma}^2} \right) dx_1. \]

- Integral may be cumbersome for long paths
- Can be simplified using the concept of Domain of Data Relevance
- See Bierlaire & Frejinger (2008) and Bierlaire, Chen and Newman (2010)
Recursion: second step

\[ \Pr(\hat{x}_1, \hat{x}_2|p) = \Pr(\hat{x}_2|\hat{x}_1, p) \Pr(\hat{x}_1|p), \]

Focus now on

\[ \Pr(\hat{x}_2|\hat{x}_1, p) = \int_{x_2 \in p} \Pr(\hat{x}_2|x_2, \hat{x}_1, p) \Pr(x_2|\hat{x}_1, p) dx_2. \]

- first term = \( \Pr(\hat{x}_2|x_2) \) measurement error, same as before
- second term: predicts the position at time \( \hat{t}_2 \) of the traveler

\[ \Pr(x_2|\hat{x}_1, p) = \int_{x_1 \in p} \Pr(x_2|x_1, \hat{x}_1, p) \Pr(x_1|\hat{x}_1, p) dx_1. \]
Position predictor

\[
\Pr(x_2|\hat{x}_1, p) = \int_{x_1 \in p} \Pr(x_2|x_1, \hat{x}_1, p) \Pr(x_1|\hat{x}_1, p) dx_1.
\]

- First term: movement model
  \[
  \Pr(x_2|x_1, \hat{x}_1, p) = f_x(x_2|x_1, \hat{t}_1, \hat{t}_2, p),
  \]

- Second term: Bayes rule
  \[
  \Pr(x_1|\hat{x}_1, p) = \frac{\Pr(\hat{x}_1|x_1, p) \Pr(x_1|p)}{\int_{x_1} \Pr(\hat{x}_1|x_1, p) \Pr(x_1|p) dx_1}.
  \]

simplifies to

\[
\Pr(x_1|\hat{x}_1, p) = \frac{\Pr(\hat{x}_1|x_1, p)}{\int_{x_1} \Pr(\hat{x}_1|x_1, p) dx_1}
\]
Measurement equations

- Step $k$ of the recursion based on same principles
- but requires some technical simplifications

$$\Pr(x_{k-1} | \hat{x}_{k-1}, p) = \frac{\Pr(\hat{x}_{k-1} | x_{k-1}, p)}{\int_x \Pr(\hat{x}_{k-1} | x, p) \, dx}.$$ 

- Integrals can be simplified using the DDR
Case study: true path
Case study: path with a deviation (1)
Case study: path with a deviation (2)
Case study: path from map matching algo

![Map with numbered points](image-url)
Case study: log likelihood from measurement equations

- True path: -11.3
- Deviation 1: -12.9
- Deviation 2: -13.2
- Map matching: $-\infty$

- DDR simplifications assign 0 probability on unrealistic paths
- Results are consistent with intuition
Stochastic map matching algorithm

- Generate a set of paths candidates
- Compute for each of them the likelihood of the GPS data
- Select a subset based on this likelihood
Summary

- We have designed a procedure that account for
  - the error of the GPS device
  - the error of the network coding
  - the movement of the smartphone
- It can involve complex models
- Technical simplifications are possible to make it operational on real data
- We have also designed
  - a path generation algorithm (Bierlaire et al. 2010)
  - a procedure for the estimation of route choice models (Bierlaire & FRejinger, 2008)
References

  
  http://dx.doi.org/10.1016/j.trc.2007.07.007


  http://transp-or.epfl.ch/php/abstract.php?type=1&id=BierChenNewm10