Generating probabilistic path observation from GPS data for route choice modelling

J. Chen J. Newman M. Bierlaire

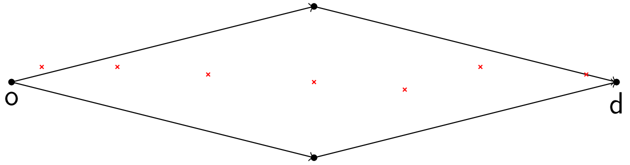
Transport and Mobility Laboratory, EPFL

European Transport Conference, 2009

Outline

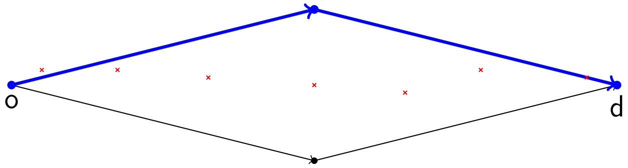
Motivation

1. Path observation need to be generated from GPS data for route choice modeling



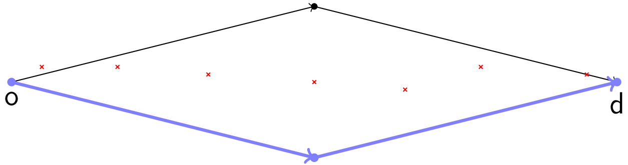
Motivation

2. Deterministic map matching algorithm may introduce biases into the observation data, if the matching is wrong.



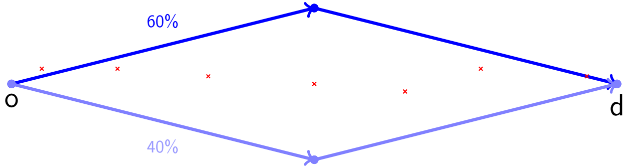
Motivation

2. Deterministic map matching algorithm may introduce biases into the observation data, if the matching is wrong.



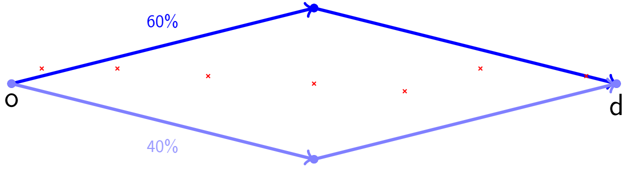
Motivation

3. Discrete route choice models accept probabilistic path representation (Bierlaire, M. and E. Frejinger 2008)



Motivation

3. Discrete route choice models accept probabilistic path representation (Bierlaire, M. and E. Frejinger 2008)



We propose a method which generates probabilistic representation of path observation

Outline

A single GPS observation

A GPS observation is recorded as $\check{g} = (\check{t}, \check{x}, \check{\sigma^x}, \check{v}, \check{\sigma^v})$, a tuple containing:

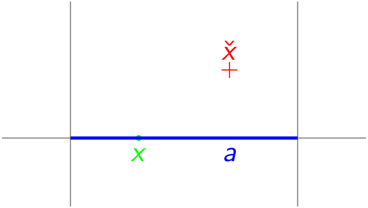
- a time stamp $\check{t}$, without error;
- a coordinates $\check{x}$, and the standard deviation of the error in the measurement of that coordinates $\check{\sigma^x}$;
- a speed measurement $\check{v}$, and the standard deviation of the error in the measurement of that speed $\check{\sigma^v}$.

Outline

Measurement for Spatial Relationships

We want to probabilistically measure the relationships:

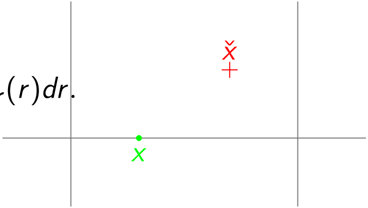
- GPS point $\check{x}$ - location x
- GPS point $\check{x}$ - arc a



Measurement for Spatial Relationships

The probability that $\check{x}$ is produced by a device at location x is given by

$$\Lambda(\check{x}, x) = \Pr(r \geq \|\check{x} - x\|) = \int_{r=\|\check{x}-x\|}^{+\infty} f_{\boldsymbol{r}}(r)dr.$$



r - horizontal error, which is distance between the true location and the observed point.

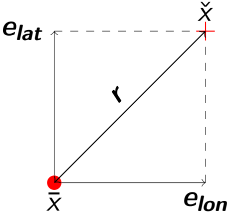
Measurement for Spatial Relationships

We assume:

- Error is independent of direction.
- $\boldsymbol{e_{lon}} \sim N(0, \sigma)$, $\boldsymbol{e_{lat}} \sim N(0, \sigma)$.

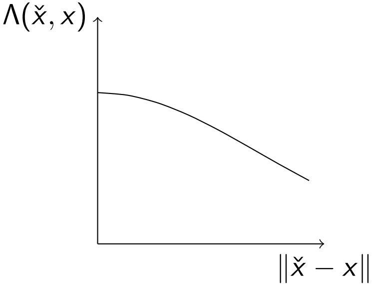
Then,

$$\boldsymbol{r} = \sqrt{\boldsymbol{e_{lon}}^2 + \boldsymbol{e_{lat}}^2} \sim Rayleigh(\sigma^{\check{x}}).$$



Measurement for Spatial Relationships

$$\Lambda(\check{x}, x) = \exp(-2\frac{\|\check{x} - x\|^2}{(\sigma^{\check{x}})^2})$$



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

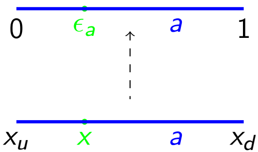
Measurement for Spatial Relationships

Representation of location in network.
ϵ_a is the position of x on arc a.

$$x = \ell_a(\epsilon_a), \qquad \epsilon_a \in [0,1]$$

ℓ_a describes the shape of arc a;

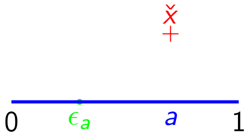
$$f_{\mathbf{g},\boldsymbol{\epsilon_a}}(\check{g}, \epsilon_a) = \Lambda(\check{x}, \ell_a(\epsilon_a)).$$



Measurement for Spatial Relationships

The probability of observing $\check{x}$ on arc a:

$$\Pr(\check{x}, a) = \int_{x \in a} \Lambda(\check{x}, x) dx$$
$$= l_a \cdot \int_0^1 \Lambda(\check{x}, \epsilon_a) d\epsilon.$$



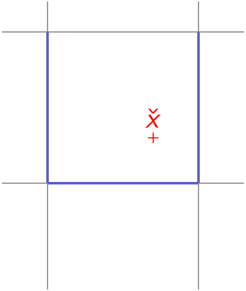
l_a is the length of arc a.

Measurement for Spatial Relationships

The probability that arc a is the true arc given observed $\check{x}$

$$\Pr(a|\check{x}) = \frac{\Pr(\check{x}, a)}{\sum_{b \in A} \Pr(\check{x}, b)}$$

Domain of relevance (D): A set of arcs which are possibly true.



TRANSP-OR

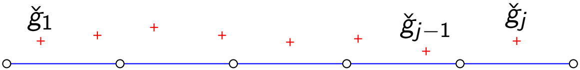
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Outline

Measurement equation for path probability

The probability of observing a GPS trace on path p:

$$\Pr(\check{g}_j, \check{g}_{j-1}, \cdots, \check{g}_1 | p) =$$



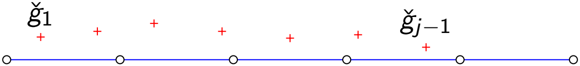
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Measurement equation for path probability

The probability of observing a GPS trace on path p:

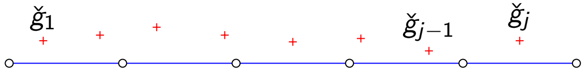
$$\Pr(\check{g}_j, \check{g}_{j-1}, \cdots, \check{g}_1 | p) = \Pr(\check{g}_{j-1}, \cdots, \check{g}_1 | p)$$



Measurement equation for path probability

The probability of observing a GPS trace on path p:

$$\Pr(\check{g}_j, \check{g}_{j-1}, \cdots, \check{g}_1 | p) = \Pr(\check{g}_{j-1}, \cdots, \check{g}_1 | p) \cdot \Pr(\check{g}_j | \check{g}_{j-1}, \cdots, \check{g}_1, p)$$



Probability of observing a GPS

Pr(observing a GPS, at a position)

$= \text{Pr(at the position)} \cdot \text{Pr(observing the GPS|at the position)}$

So,

$$\Pr(\breve{g}_j|\breve{g}_{j-1},\cdots,\breve{g}_1,p) = \sum_{a\in(D_j\cap p)} \Pr(\breve{g}_j, a|\breve{g}_{j-1},\cdots,\breve{g}_1,p)$$

$$= \sum_{a\in(D_j\cap p)} l_a \cdot \int_0^1 f_{\epsilon^j}(\epsilon_a|\breve{g}_{j-1},\cdots,\breve{g}_1,p) \cdot f_{\mathbf{g},\varepsilon^j}(\breve{g}_j,\epsilon_a)d\epsilon_a$$

Outline

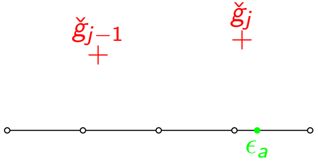
State function $f_{\boldsymbol{\epsilon} j}(\epsilon_a | \check{g}_{j-1}, \cdots, \check{g}_1, p)$

- State function is the PDF of the current state (position ϵ_a), given 1. the GPS trace before the current observation, and 2. the current timestamp.
- The position ϵ_a where $\check{g}_j$ is observed, depends on
  1. the position where $\check{g}_{j-1}$ is observed,
  2. traveler's movement during $[\check{t}_{j-1}, \check{t}_j]$.

Movement between GPS observations

Travel from previous GPS

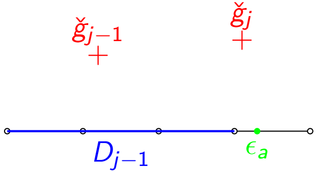
$$f_{\epsilon^j}(\epsilon_a|\check{g}_{j-1},\cdots,\check{g}_1,p)$$
$$=$$



Movement between GPS observations

Travel from previous domain

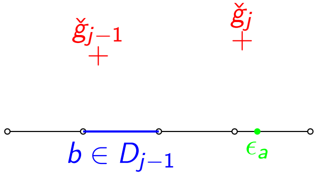
$$
\begin{aligned}
& f_{\epsilon^j}(\epsilon_a | \check{g}_{j-1}, \cdots, \check{g}_1, p) \\
= & f_{\epsilon^j}(\epsilon_a | D_{j-1}) \cdot \\
& \Pr(D_{j-1} | \check{g}_{j-1}, \cdots, \check{g}_1, p)
\end{aligned}
$$



Movement between GPS observations

Travel from previous arc

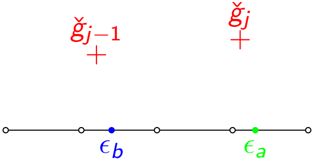
$$
\begin{aligned}
& f_{\epsilon^j}(\epsilon_a | \breve{g}_{j-1}, \cdots, \breve{g}_1, p) \\
= & \sum_{b \in D_{j-1}} f_{\epsilon^j}(\epsilon_a | b) \cdot \\
& \Pr(b | D_{j-1}, \breve{g}_{j-1}, \cdots, \breve{g}_1, p) \cdot \\
& \Pr(D_{j-1} | \breve{g}_{j-1}, \cdots, \breve{g}_1, p)
\end{aligned}
$$



Movement between GPS observations

Travel from previous position

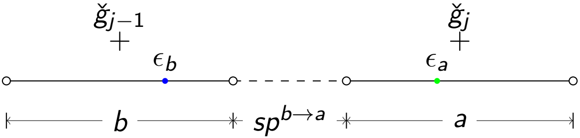
$$
\begin{aligned}
& f_{\epsilon^j}(\epsilon_a|\check{g}_{j-1},\cdots,\check{g}_1,p) \\
= & \sum_{b\in D_{j-1}} \int_{\epsilon_b=0}^{1} f_{\epsilon^j}(\epsilon_a|\epsilon_b)\cdot \\
& f_{\epsilon^{j-1}}(\epsilon_b|b)d\epsilon_b\cdot \\
& \Pr(b|D_{j-1},\check{g}_{j-1},\cdots,\check{g}_1,p)\cdot \\
& \Pr(D_{j-1}|\check{g}_{j-1},\cdots,\check{g}_1,p)
\end{aligned}
$$



Outline

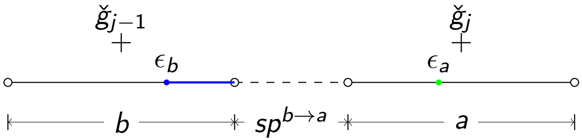
State transition between adjacent GPS points

$\boldsymbol{t}_{b \to a} =$



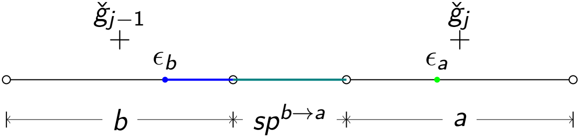
State transition between adjacent GPS points

$$\boldsymbol{t}_{b\rightarrow a} = (1-\epsilon_b)\cdot \boldsymbol{t}_b$$



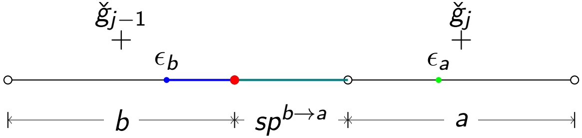
State transition between adjacent GPS points

$$\boldsymbol{t}_{b\rightarrow a} = (1-\epsilon_b)\cdot \boldsymbol{t}_b + \sum_{c\in sp^{b\rightarrow a}} \boldsymbol{t}_c$$



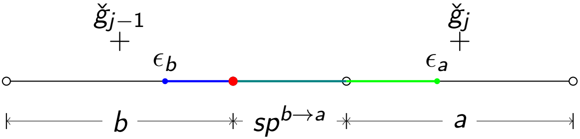
State transition between adjacent GPS points

$$\boldsymbol{t}_{b\to a} = (1-\epsilon_b)\cdot \boldsymbol{t}_b + \sum_{c\in sp^{b\to a}} \boldsymbol{t}_c + \boldsymbol{t}_w^{b\to a}$$



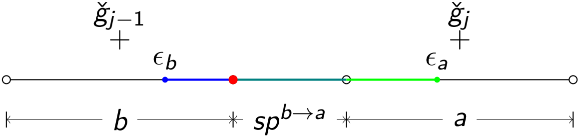
State transition between adjacent GPS points

$$\boldsymbol{t}_{b \to a} = (1 - \epsilon_b) \cdot \boldsymbol{t}_b + \sum_{c \in sp^{b \to a}} \boldsymbol{t}_c + \boldsymbol{t}_w^{b \to a} + \epsilon_a \cdot \boldsymbol{t}_a$$



State transition between adjacent GPS points

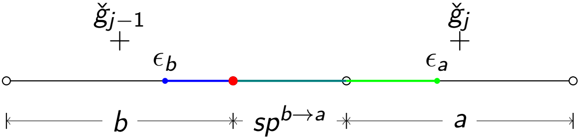
$$\boldsymbol{t}_{b\to a} = (1-\epsilon_b)\cdot \boldsymbol{t}_b + \sum_{c\in sp^{b\to a}} \boldsymbol{t}_c + \boldsymbol{t}_w^{b\to a} + \epsilon_a \cdot \boldsymbol{t}_a = \check{t}_j - \check{t}_{j-1}$$



State transition between adjacent GPS points

$$f_{\boldsymbol{\epsilon}^j}(\epsilon_a|\epsilon_b) = f_{\boldsymbol{t}_{b\rightarrow a}}(\check{t}_j - \check{t}_{j-1})$$

$$\boldsymbol{t}_{b\rightarrow a} = (1-\epsilon_b)\cdot \boldsymbol{t}_b + \sum_{c\in sp^{b\rightarrow a}} \boldsymbol{t}_c + \boldsymbol{t}_w^{b\rightarrow a} + \epsilon_a \cdot \boldsymbol{t}_a = \check{t}_j - \check{t}_{j-1}$$



Travel time on arc a and b

The speed is assumed to be constant during the traveler travels on an arc, then travel time on arc is calculated by

$$t_c = \frac{l_c}{v_c}, \qquad \forall c \in A.$$

On arc a and b, speeds are observed with errors. We use normal distribution to estimate the speeds,

$$\boldsymbol{v_b} \sim N(\check{v}_{j-1}, \check{\sigma}^{v}{}_{j-1})$$
$$\boldsymbol{v_a} \sim N(\check{v}_{j}, \check{\sigma}^{v}{}_{j})$$

Travel time on arc $c \in sp^{b \to a}$

In traffic theory, the free flow speed ratio reflects the traffic conditions. The inverse free flow speed ratio is

$$\varpi = \bar{v}/v$$

Travel time on arc $c \in sp^{b \to a}$

In traffic theory, the free flow speed ratio reflects the traffic conditions. The inverse free flow speed ratio is

$$\varpi = \bar{v}/v$$

At each GPS observation, the inverse free flow speed ratio is calculated by

$$\varpi_j = \sum_{a \in D_j} \Pr(a|\breve{g}_j) \cdot \frac{\bar{v}_a}{\check{v}_j}.$$

Travel time on arc $c \in sp^{b \to a}$

We assume that within a certain geographical area and time period, the traffic condition is stable to some extent, then

$$\boldsymbol{\varpi} \sim N(\bar{\varpi}, \delta_{\varpi})$$

Travel time on arc $c \in sp^{b \to a}$

We assume that within a certain geographical area and time period, the traffic condition is stable to some extent, then

$$\boldsymbol{\varpi} \sim N(\bar{\varpi}, \delta_{\varpi})$$

$$\bar{\varpi} = \frac{1}{n} \sum_{\check{g}_i \in \Theta_j} \varpi_i \qquad , \delta_{\varpi} = \sqrt{\frac{1}{n-1} \sum_{\check{g}_i \in \Theta_j} (\varpi_i - \bar{\varpi})^2}.$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Travel time on arc $c \in sp^{b \to a}$

We assume that within a certain geographical area and time period, the traffic condition is stable to some extent, then

$$\boldsymbol{\varpi} \sim N(\bar{\varpi}, \delta_{\varpi})$$

$$\bar{\varpi} = \frac{1}{n} \sum_{\check{g}_i \in \Theta_j} \varpi_i \qquad , \delta_{\varpi} = \sqrt{\frac{1}{n-1} \sum_{\check{g}_i \in \Theta_j} (\varpi_i - \bar{\varpi})^2}.$$

For each $c \in sp^{b \to a}$,

$$\boldsymbol{t}_c = \frac{l_c}{\bar{v}_c} \cdot \boldsymbol{\varpi}$$

follows normal distribution.

Waiting time caused by stops

- Stops caused by traffic control devices.

Waiting time caused by stops

- Stops caused by traffic control devices.
- Probability that an observation is recorded during waiting is high.

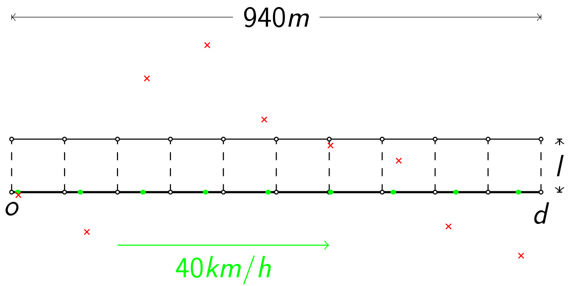
Waiting time caused by stops

- Stops caused by traffic control devices.
- Probability that an observation is recorded during waiting is high.
- Waiting time is a penalty to unlikely state transitions.

Waiting time caused by stops

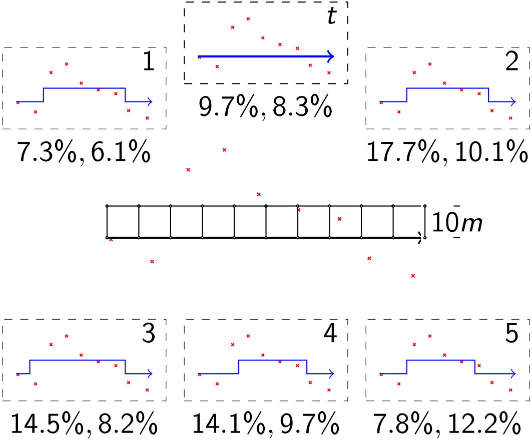
- Stops caused by traffic control devices.
- Probability that an observation is recorded during waiting is high.
- Waiting time is a penalty to unlikely state transitions.
- Distribution is assumed to be uniform.

Simulation scenario

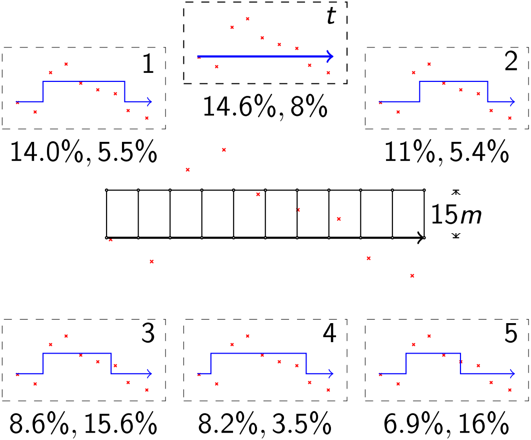


We run the algorithm with l being $10m$, $15m$, $20m$ respectively.

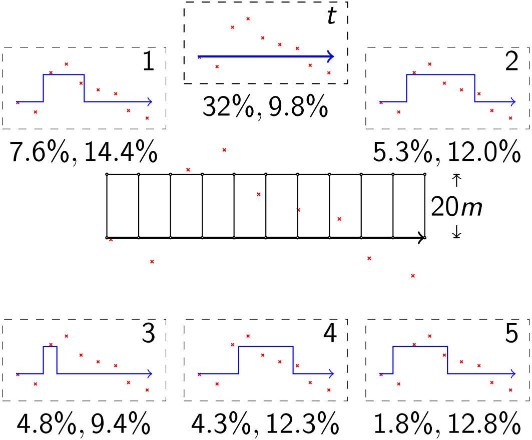
Result with $l = 10m$



Result with $l = 15m$



Result with $l = 20m$



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Future work

- Model the situation of stops (GPS observations with very low speed).
- Run algorithm on real data.
- Compare against map-matching algorithms.
- Estimate route choice models.