Detecting pedestrian destinations from ubiquitous digital footprints

Antonin Danalet
TRANSP-OR, EPFL

FCL-Talk
January 22\textsuperscript{nd}, 2013
Transport and Mobility lab

- People: Prof. Michel Bierlaire, 4 postdocs, 9 PhD students
- Research streams:
  - Transportation Research,
  - Operations Research,
  - Discrete Choice Models
- 1 multidisciplinary subgroup: pedestrian dynamics
- 2 projects related to pedestrians:
  - Swiss National Science Foundation
  - Swiss Railways (SBB-CFF-FFS)
## Image in the (Swiss) population

### 4. Perceptions des modes de transport

Pour chacun des moyens de transport suivants, indiquez 3 adjectifs qui, selon vous, les décrivent le mieux.

<table>
<thead>
<tr>
<th></th>
<th>Adjectif 1</th>
<th>Adjectif 2</th>
<th>Adjectif 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>La VOITURE est :</td>
<td>Comfortable</td>
<td>Convenient</td>
</tr>
<tr>
<td>2</td>
<td>Le TRAIN est :</td>
<td>Comfortable</td>
<td>Fast</td>
</tr>
<tr>
<td>3</td>
<td>Le BUS, METRO et TRAM sont :</td>
<td>Convenient</td>
<td>Fast</td>
</tr>
<tr>
<td>4</td>
<td>Le CAR POSTAL est :</td>
<td>Comfortable</td>
<td>Convenient</td>
</tr>
<tr>
<td>5</td>
<td>Le VELO est :</td>
<td>Healthy</td>
<td>Sporty</td>
</tr>
<tr>
<td>6</td>
<td>MARCHER est :</td>
<td>Healthy</td>
<td>Slow</td>
</tr>
</tbody>
</table>

Source: Optima, Projet de recherche sur la mobilité combinée :
Rapport définitif de l’enquête de préférences révélées, EPFL
http://transport.epfl.ch/optima
Challenge

Infrastructure

Behavior
Presentation outline

- Activity-based modeling: From car to pedestrians
- The importance of data: Nur was gezählt wird zählt
- A Bayesian approach to mix map data, WiFi traces and (train or class) schedules: A case study on campus
- Toward an activity-based model for pedestrians How and what can we learn?
Activity-based modeling: From car to pedestrians
Travel demand modeling: 1960s

- Rapid increase in car ownership and usage
- Need to assess the impact of investments
- Aggregate, gravity type models
- **4-step model**
  - trip generation
  - trip distribution
  - mode choice
  - route choice
Travel demand modeling: 1970s

Shifting paradigm in travel demand modeling (1): from gravity to discrete choice
- direct modeling of individual choice behavior, and so more sensitive to policy
- require smaller data sets for calibration
- incorporate more explanatory variables
Travel demand modeling: 1980s

Shifting paradigm in travel demand modeling (2): travel behavior as a derivative of activities

- emphasize on activity scheduling behavior
- compliant with more complex patterns
- compliant with new policies
  (e.g., congestion pricing, ridesharing, ...)

A. Danalet

TRANSP-OR ENAC EPFL
Hägerstrand’s Time Geography
Research on pedestrians

- 3 levels of pedestrian behavior
  - Strategical (destination, activity choices)
  - Tactical (route choice)
  - Operational (walking behavior)
- First developments in mid-90’s
- Still a new research area, with a focus on operational level
- TRANSP-OR active in this area, e.g., walking behavior as a choice of next step
Model

Activity 1
• Activity-episode 1

Activity 2
• Activity-episode 2

Activity 3
• Activity-episode 3
Fachtagung 2010, Fussverkehr Schweiz, Fachverband der FussgängerInnen (workshop of the Swiss Association of Pedestrians)

(only what you can count counts)

**NUR WAS GEZÄHLT WIRD ZÄHLT**
Data: from a device-centric approach...
Data: ... to infrastructure traces
Paléo
Paléo
Paléo
Bluetooth traces from smartphones with GPS

Video not available in PDF format
Please visit:
http://www.youtube.com/watch?v=8Zi45m67jbE

Video: Julien Eberle
http://people.epfl.ch/julien.eberle

Data: F. M. Naini et al., Population Size Estimation Using a Few Individuals as Agents
http://infoscience.epfl.ch/record/169801/files/MovDTV11.pdf
Infrastructure traces: strengths

- Data related to **infrastructure**, not to individuals
- Full coverage of the facility is **cheap** and allows for estimating the **overall demand**
- Infrastructure partially already **exists**, and increasing it has **positive side effects**
WiFi Traces on EPFL Campus

- 789 access points
- 2 datasets:
  - AP to which you are connected
  - Triangulation data (Cisco)

Pros
- Already available
- Covering the full infrastructure

Cons
- Low precision
Pedestrian network
Pedestrian network

- 4 levels of path (major, inter-, intra-building, access to rooms)
- 56'655 edges, 50'131 vertices
- 17'502 public “points of interest”
- 13’783 “rooms”
Class schedules at EPFL

Number of students following courses in Spring 2012 at EPFL (schedules, not counting data)
A PROBABILISTIC METHOD FOR ESTIMATING PEDESTRIAN ACTIVITY-EPISODES SEQUENCES

A. Danalet et al., Estimating Pedestrian Destinations using Traces from WiFi Infrastructures
http://infoscience.epfl.ch/record/180079
**Goal:** extract the possible activity-episodes performed by pedestrians from digital traces from communication networks

**Input**
- measurement
- prior
- pedestrian map

**Output**
- set of candidate activity-episodes sequences associated with the probability of being the true one
Probabilistic method

- Probabilistic measurement model
- Generation of activity-episode sequences
- Intermediary signals
- Sequence elimination procedure
Definitions

- Measurement: \( \hat{s} = (\hat{x}, \hat{t}) \)
- Activity-episode: \( a = (x, t^-, t^+) \)
- Episode location, start time and end time
- Activity-episode sequence: \( (a_1, \ldots, a_m) = a_{1:m} \)
- Activity: \( A(a) \)
- Activity pattern: \( (A_1, \ldots, A_m) = A_{1:m} \)
Probabilistic measurement model

\[ P(a_1:m \mid \hat{s}_{1:n}) = \frac{P(\hat{s}_{1:n} \mid a_1:m) \cdot P(a_1:m)}{\sum_{a \in A} P(\hat{s}_{1:n} \mid a_1:m) \cdot P(a_1:m)} \]

- Measurement likelihood
- Prior
- Activity model
Probabilistic measurement model

\[
P(\hat{s}_{1:n}|a_{1:m}) = \prod_{j=1}^{m} P(\hat{s}_{i_{j-1}+1:i_{j}}|a_{j}) \]

\[
= \prod_{j=1}^{m} \prod_{i=1}^{n} P(\hat{s}_{i,j}|a_{j}) \]

\[
= \prod_{j=1}^{m} \prod_{i=1}^{n} P(\hat{x}_{i,j}|x_{j}) \]
Prior

\[ P(x, t^-, t^+) \]

\[ P(\text{classroom}) = P(\text{offices}) = P(\text{restaurants}) = P(\text{others}) = \frac{1}{4} \]

\[ P(x|\text{classroom}) = \frac{\int_{t=t^-}^{t^+} f_x(t)\,dt}{\int_{t=t^-}^{t^+} f(t)\,dt} \]
Generation of activity-episode sequences
Generation of activity-episode sequences
Generation of activity-episode sequences

$t_{i+1}^- \sim U(t_i^+, t^+_{i+1} - t_{x_i,x_{i+1}})$

$t_{i+1}^+ \sim U(t_i^- + t_{x_i,x_{i+1}}, \hat{t}_{i+1})$
Generation of activity-episode sequences

\[ f(t_{i+1}) = \frac{1}{\hat{t}_{i+1} - tt_{x_i,x_{i+1}} - \hat{t}_i} \ln \frac{\hat{t}_{i+1} - tt_{x_i,x_{i+1}} - \hat{t}_i}{\hat{t}_{i+1} - t_{i+1}} \]
Intermediary signals

- Eliminate intermediary signal if

\[ E(t^+) - E(t^-) < T_{\text{min}} \]

since we generate an activity episode at each signal.
Sequence elimination

\[ E(t) = E(t) - E(t) < T_{\text{min}} \]

Diagram:

- Root node
- Node 1: \( a_i^1 \)
  - \( a_{i+1}^1 \)
  - \( a_{i+1}^2 \)
- Node 2: \( a_i^2 \)
  - \( a_{i+1}^1 \)
  - \( a_{i+1}^2 \)
- Node 3: \( a_i^3 \)
  - \( a_{i+1}^1 \)
  - \( a_{i+1}^2 \)

Values:

- \( a_i^1: 0.1 \)
- \( a_i^2: 0.05 \)
- \( a_i^3: 0.3 \)
- \( a_{i+1}^1: 0.2 \)
- \( a_{i+1}^2: 0.15 \)
- \( a_{i+1}^3: 0.2 \)
WiFi Traces on EPFL Campus: Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Time spent</th>
<th>Floor</th>
<th>Location</th>
<th>Time spent</th>
<th>Floor</th>
<th>Location</th>
<th>Δx</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$U(8.40, 8.40) - U(8.40, 9.31)$</td>
<td>0</td>
<td>Classroom</td>
<td>8.32-10.30</td>
<td>1</td>
<td>Classroom</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>$U(8.42, 9.33) - U(10.38, 10.38)$</td>
<td>1</td>
<td>Classroom</td>
<td>8.32-10.30</td>
<td>1</td>
<td>Classroom</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>$U(10.38, 11.18) - U(11.51, 11.51)$</td>
<td>3</td>
<td>Office</td>
<td>Until 11.47</td>
<td>3</td>
<td>Author’s office</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>$U(11.52, 12.00) - U(12.47, 12.47)$</td>
<td>2</td>
<td>Classroom</td>
<td>From 11.55</td>
<td>1</td>
<td>Restaurant</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>$U(12.48, 13.03) - U(13.03, 13.44)$</td>
<td>3</td>
<td>Office</td>
<td>Around 13.00</td>
<td>3</td>
<td>Author’s office</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>$U(13.06, 13.47) - U(13.53, 13.53)$</td>
<td>2</td>
<td>Restaurant</td>
<td>Around 14.00</td>
<td>2</td>
<td>Cafeteria</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>$U(13.53, 14.10) - U(19.40, 19.42)$</td>
<td>3</td>
<td>Office</td>
<td>Until around 19.45</td>
<td>3</td>
<td>Author’s office</td>
<td>8</td>
</tr>
</tbody>
</table>
WiFi Traces on EPFL Campus: Results
WiFi Traces on EPFL Campus: Results

Legend
Activities
Weighted shortest path
Pedestrian network
WiFi Traces on EPFL Campus: Results

Video not available in PDF format
Please visit:
http://www.youtube.com/watch?v=SEp-yNXLfUY
How can we learn about activities?

- WiFi infrastructure:
  - Cheap
  - Positive side effects
  - Dense and covers the whole station

- Map knowledge
  - Compensate weakness of localization

- Privacy
  - Data already exist
  - Localization is weak
  - Daily anonymization
What’s next?

- Socioeconomic data:
  - Survey?
  - Heterogenous network?
- Develop a location choice model
  - Choice set generation
  - Measurement equation for observations
  - Dynamic of the system
  - Specification of the model
- Move from location choice to activity choice…
What can we learn about activities?

- Overall demand (with calibration)
  - How many people are going on platforms and in shops
- Understanding underlying reasons for activity behavior
  - Why people (don’t) go in this shop
- Forecasting demand for activities in case of different scenarios
  - What would happen if this type of shop is moved
Slides and contact information:
http://people.epfl.ch/antonin.danalet