A Bayesian Approach to Detect Pedestrian Destination-Sequences from WiFi Signatures

Antonin Danalet
Bilal Farooq
Michel Bierlaire

TRISTAN VIII, San Pedro de Atacama
June 10th, 2013
Presentation outline

- Motivation
- Data requirement
- Methodology
- A case study on EPFL campus
- Conclusion
- Future work
MOTIVATION
Walking is the key for efficient multimodal transport systems

Crowd in a railway station in Mumbai, India
Photo: National Geographic
Lake Geneva region

By 2030, 100’000 passengers per day between Geneva and Lausanne

- 2000 travelers/day between Geneva and Lausanne
- > 25’000 travelers/day between Geneva and Lausanne
- > 50’000 travelers/day between Geneva and Lausanne
- > 100’000 travelers/day between Geneva and Lausanne

* Forecast by Swiss Railways for the maximum scenario
Understand pedestrian activities

What we are doing: Campus

What we want to do: Station
Challenges

- Detect pedestrian destinations
- Model pedestrian activity scheduling behavior
- Forecast the impact of changes in the infrastructure

Carlstein, T. (1978)
DATA REQUIREMENT
Data requirement

- **Required**
  - Localization data with full coverage of the facility
  - Semantically-enriched routing graph for pedestrians

- **Not really required but often available information**
  - Prior potential attractiveness
Data requirement: Localization

- Data from communication network infrastructure
  - GSM traces (Calabrese et al. 2011, Bekhor et al. 2011)
  - WiFi traces
- Data processing is needed (Rieser-Schüsseler 2012)
  - Detection of stop points
  - Activity purpose detection through land-use information and spatial matching
Data requirement: Pedestrian network

- We need maps
  - With **points of interests** (space)
  - With **shortest path** (time)
- More and more available in airports, malls, museums, campuses, hospitals
  - **Nokia**: 214 shopping malls in 2011, 4605 indoor maps in July 2012, 5100 in December 2012
  - **Microsoft**: 2700 indoor maps
  - **Google**: > 10’000 indoor maps
  - **Start-ups**: Wifarer, Meridian, Point Insider, ByteLight
Data requirement: Potential attractivity

- **Potential attractivity** $C(x,t)$ depends on
  - destination $x$
    - Classroom, platform, scene, …
  - time $t$
    - class schedules, train schedules, opening hours, …

- **Examples**:
  - 1500 passengers on platform 4 arriving at 16h04
  - 32 students in a classroom from 8h15 to 10h
  - 400 seats in a restaurant open from 11h to 14h30
METHODOLOGY
**Methodology**

**Goal:** extract the possible activity-episodes performed by pedestrians from digital traces from communication networks

- Localization measurement
- Potential attractivity
- Semantically-enriched routing graph

**Output**
- set of candidate activity-episodes sequences associated with the likelihood to be the true one
Definitions / Notations

- 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: \( (A_1, \ldots, A_m) = A_{1:m} \)
Methodology

- Probabilistic measurement model: A Bayesian approach
  - Measurement equation
  - Prior

- Generation of activity-episode sequences
  - Episode location
  - Episode start and end times
Probabilistic measurement model

Measurement likelihood \( P(a_{1:m} | \hat{s}_{1:n}) \propto P(\hat{s}_{1:n} | a_{1:m}) \cdot P(a_{1:m}) \)

Prior

Activity model
Measurement error

\[
P(\hat{s}_{1:n} | a_{1:m}) = \prod_{j=1}^{m} P(\hat{s}_{i_{j-1}+1:i_j} | a_j) \quad \text{Independence between activities}
\]

\[
= \prod_{j=1}^{m} \prod_{i=1}^{n} P(\hat{s}_{i_j} | a_j) \quad \text{Independence between signals}
\]

\[
= \prod_{j=1}^{m} \prod_{i=1}^{n} P(\hat{x}_{i_j} | x_j) \quad \text{No time measurement error}
\]
Prior

\[
P(a_{1:m}) = \prod_{j=1}^{m} P(a_j) \quad (1)
\]

\[
= \prod_{j=1}^{m} P(x_j, t_j^-, t_j^+) \quad (2)
\]

\[
= \prod_{j=1}^{m} \frac{C_{x_j}(t_j^-, t_j^+)}{\sum_{x \in X} C_x(t_j^-, t_j^+)} \quad (3)
\]

\[
= \prod_{j=1}^{m} \int_{t_j = t_j^-}^{t_j^+} \frac{C_{x_j}(t_j)}{\sum_{x \in X} C_x(t_j)} dt \quad (4)
\]
Generation of activity-episode sequences
Generation of activity-episode sequences

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

Diagram:
- Root node
- Subnodes:
  - $a_i^1$
  - $a_i^2$
  - $a_i^3$
  - $a_{i+1}^1$
  - $a_{i+1}^2$
  - $a_{i+1}^1$
  - $a_{i+1}^1$
  - $a_{i+1}^2$
  - $a_{i+1}^2$
Intermediary signals

- Eliminate intermediary signal if

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

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

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

\[
\begin{align*}
  a_i^1 & \quad a_i^2 \quad a_i^3 \\
  a_{i+1}^1 & \quad a_{i+1}^2 \quad a_{i+1}^3 \\
  0.1 & \quad 0.05 & \quad 0.3 & \quad 0.2 & \quad 0.15 & \quad 0.2
\end{align*}
\]
A CASE STUDY ON EPFL CAMPUS
EPFL data: Localization

- 8 participants for 2 months with known ID
- Non-participants: 46 days, but only 10 with courses
  - 200 students in 6 different classes
  - 317 employees
  - 700 students from University of Lausanne
- For 151 CE students, 152’598 observations
- Precision: 191 meters
EPFL data: Pedestrian network

- Source: map.epfl.ch
- 56’655 edges
- 4 different levels of path
  - Major (« highway »)
  - Inter-building
  - Intra-building
  - Access to offices
- Shortest path
- All offices, restaurants, classrooms and other points of interest are coded: X
EPFL data: Potential attractiveness $C(x,t)$

- Class schedules with
  - Number of students
  - Name of the classroom
- Number of employees per office
  - Name of the office
  - Sum of percent of work (e.g., 3 full times = 300%)
- Number of seats in restaurants
  - Localization
  - Opening hours
- Number of seats in library
### Results

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arrival time</td>
<td>Departure time</td>
<td>Floor</td>
<td>Location</td>
<td>Time spent</td>
<td>Floor</td>
<td>Location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8:33-8:33</td>
<td>10:38-10:38</td>
<td>1</td>
<td>Classroom</td>
<td>8.32am-10.30am</td>
<td>1</td>
<td>Classroom</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:40-10:40</td>
<td>11:51-11:51</td>
<td>3</td>
<td>Office</td>
<td>Until 11.47am</td>
<td>3</td>
<td>Author’s office</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:54-11:54</td>
<td>12:47-12:53</td>
<td>1</td>
<td>Restaurant</td>
<td>From 11.55 am</td>
<td>1</td>
<td>Restaurant</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:51-12:58</td>
<td>13:03-13:44</td>
<td>3</td>
<td>Office</td>
<td>Around 1pm</td>
<td>3</td>
<td>Author’s office</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13:06-13:47</td>
<td>13:53-14:02</td>
<td>2</td>
<td>Cafeteria</td>
<td>Around 2pm</td>
<td>2</td>
<td>Cafeteria</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13:55-14:04</td>
<td>19:45-19:45</td>
<td>3</td>
<td>Office</td>
<td>Until around 7.45pm</td>
<td>3</td>
<td>Author’s office</td>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SENSITIVITY ANALYSIS
Sensitivity analysis: prior

● With flat prior
  - # destinations / Start and end time: OK
  - Distance / category of destination: Not OK
● Attractivity of visited destinations should be 3x bigger than of non-visited destinations
● Global capacity creates bias
CONCLUSION
Conclusion

- Prior needed to **overcome low precision**
- **Localization data brings dynamics** in the model
- Pedestrian map gives:
  - Spatial information
  - Temporal information
- Our methodology is **merging** these different types of data
- Robust for **low density data**
FUTURE WORK
Future work

- Binary choice model for attendance of scheduled activity
- Actual start and end times of scheduled activity-episodes
- Analysis of the access to the facility
  - First and last destination of the sequence
  - Arrival times, departure times
- Based on class attendance and on available time budget: activity scheduling
THANK YOU
References

