Modeling the dynamics of all-day activity plans

Willem Himpe$^1$
Gunnar Flötteröd, Ricardo Hurtubia, and Michel Bierlaire$^2$

June 22, 2010

$^1$KU Leuven
$^2$EPFL Lausanne
Motivation

- activity-based travel demand modeling
- personalized services on smartphones
Outline

Methodology

Model estimation

Preliminary validation

Summary, outlook
Outline

Methodology

Model estimation

Preliminary validation

Summary, outlook
**Terminology**

**activity type:** home, work, education, leisure, shopping, other

**activity schedule:** temporal sequence of activity types

<table>
<thead>
<tr>
<th>home</th>
<th>work</th>
<th>shop</th>
<th>leisure</th>
<th>home</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:00</td>
<td>08:00</td>
<td>17:00</td>
<td>19:00</td>
<td>22:00</td>
</tr>
</tbody>
</table>
Dynamics of activity schedules

• timing
  – facility (e.g., shop, office) opening times
  – avoid exhaustive activities early in the day

• duration
  – being at work for 8 h is desirable, just 2 h are not
  – playing tennis for 2 h is fun, playing for 8 h is not

• sequencing
  – don’t go shopping twice per day
  – bring kids to kindergarten → pick them up later
Probabilistic activity sequencing model

- activity schedule consists of $N$ activity slots with fixed timing
- $n$th activity is $a_n \in \{\text{home, work, edu, leisure, shop, other}\}$
- $M$ unknown activities with indices $x_m$, $m = 1 \ldots M$
- objective: model probability distribution of full schedule

$$P(a_{\{x_1 \ldots x_M\}}|a_{\{1 \ldots N\}\{x_1 \ldots x_M\}})$$
Single gap model

- problem with full model: combinatorial explosion
- assume that only one activity is unknown

\[ P(a_x|a_{\{1...N\}\setminus x}) \]

- one-dimensional distribution, no combinatorial issues
- can be estimated from data
Multiple-gap model

• reconstruct full (multiple-gap) distribution

\[ P(a_{x_1...x_M} | a_{1...N}\setminus\{x_1...x_M\}) \]

from marginals

\[ P(a_{x_m} | a_{1...N}\setminus x_m), \ m = 1...M \]

• computational technique: Gibbs sampling
• only requires the estimation of single-gap models
Outline

Methodology

Model estimation

Preliminary validation

Summary, outlook
Data

- Swiss microcensus 2005
  - overall 33,390 respondents
  - activity and travel behavior for a single day
  - linked to socio-economics of respondents

- consider only canton of Vaud
  - 2157 persons
  - 8508 activities
Study region: canton of Vaud
Model structure

- represent single-gap model through decision tree
- example:
# Explanatory variables

<table>
<thead>
<tr>
<th>name</th>
<th>values</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>start_day</code></td>
<td>minutes after midnight</td>
<td>first time at which home is left</td>
</tr>
<tr>
<td><code>activ_start</code></td>
<td>minutes after midnight</td>
<td>starting time of the activity gap</td>
</tr>
<tr>
<td><code>duration</code></td>
<td>minutes</td>
<td>duration of the activity gap</td>
</tr>
<tr>
<td><code>{tot_activity}</code></td>
<td>minutes</td>
<td>total duration of activity outside gap</td>
</tr>
<tr>
<td><code>employment</code></td>
<td>full time, part time, student, unemployed</td>
<td>daytime occupation of respondent</td>
</tr>
<tr>
<td><code>weekday</code></td>
<td>any weekday</td>
<td>considered day of the week</td>
</tr>
<tr>
<td><code>prev_act</code></td>
<td>an activity type</td>
<td>activity conducted before the gap</td>
</tr>
<tr>
<td><code>next_act</code></td>
<td>an activity type</td>
<td>activity conducted after the gap</td>
</tr>
</tbody>
</table>
Estimation of single-gap model

- use C4.5 learning algorithm (Weka software package)
- incrementally builds a decision tree, then extracts rules
- selects branching conditions that separate the data well
- technically, maximizes entropy of child node distribution
- yields “crisp rules”, no probabilities
  ⇒ attach empirical activity distribution to each leaf
Estimation results

• tree with 57 leaves
• > 68% of data correctly classified (10-fold cross-validation)
• extracted rules are plausible
• example:

\[
\text{IF } \text{act\_start\_time} \leq 532 \text{ min (08:52)} \\
\text{AND employment } \notin \{\text{student, unemployed}\} \\
\text{THEN } Pr(\text{gap=work}) = 0.81
\]
## Estimation results

<table>
<thead>
<tr>
<th>work</th>
<th>shop</th>
<th>leisure</th>
<th>home</th>
<th>edu</th>
<th>other</th>
<th>model/data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1308</td>
<td>57</td>
<td>218</td>
<td>127</td>
<td>22</td>
<td>0</td>
<td>work</td>
</tr>
<tr>
<td>89</td>
<td>568</td>
<td>486</td>
<td>81</td>
<td>5</td>
<td>6</td>
<td>shop</td>
</tr>
<tr>
<td>162</td>
<td>293</td>
<td>2251</td>
<td>295</td>
<td>31</td>
<td>2</td>
<td>leisure</td>
</tr>
<tr>
<td>114</td>
<td>59</td>
<td>259</td>
<td>1277</td>
<td>4</td>
<td>1</td>
<td>home</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>47</td>
<td>15</td>
<td>391</td>
<td>0</td>
<td>education</td>
</tr>
<tr>
<td>32</td>
<td>124</td>
<td>133</td>
<td>23</td>
<td>3</td>
<td>9</td>
<td>other</td>
</tr>
</tbody>
</table>
Outline

Methodology

Model estimation

Preliminary validation

Summary, outlook
Smartphone data collection campaign

- joint project with Nokia
- approx. 50 respondents carry smartphones that observe
  - GPS, wireless, Bluetooth
  - actions conducted on phone
  - visual & acoustic samples

- supplementary survey
  $\Rightarrow$ behavioral modeling

- preliminary data base
  - one respondent
  - 45 days
Results – performance measures

• single-gap log-likelihood

\[
\mathcal{L}_{\text{single}} = \sum_{d=1}^{45} \sum_{x=1}^{N_d} \log P(a_x = y_{dx} | a_{\{1...N_d\} \setminus x})
\]

where

- \( N_d \) is the number of activities in day \( d \)
- \( y_{dx} \) is the reported activity in slot \( x \) of day \( d \)

• multiple-gap log-likelihood

\[
\mathcal{L}_{\text{multiple}} = \sum_{d=1}^{45} \sum_{x=1}^{N_d} \log P(a_x = y_{dx} | \cdot)
\]
Results – performance measures

- null log-likelihood

\[ \mathcal{L}_0 = \sum_{d=1}^{45} \sum_{x=1}^{N_d} \log \frac{1}{6} \]

- results

<table>
<thead>
<tr>
<th></th>
<th>( \mathcal{L}_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathcal{L}_0 )</td>
<td>-49.802</td>
</tr>
<tr>
<td>( \mathcal{L}_{\text{single}} )</td>
<td>-28.434</td>
</tr>
<tr>
<td>( \mathcal{L}_{\text{multiple}} )</td>
<td>-36.224</td>
</tr>
</tbody>
</table>
Outline

Methodology

Model estimation

Preliminary validation

Summary, outlook
Summary

- dynamic, non-behavioral model of activity scheduling
- requires estimation only of models for single activities
- draws full activity schedules with Gibbs sampling
Outlook

- replace known activities by activity distributions
- model more degrees of freedom
  - temporal structure
  - number of activities
  - travel episodes
- replace rule set by behavioral model
  - interpretability
  - extrapolation
Outlook

• replace known activities by activity distributions
• model more degrees of freedom
  – temporal structure
  – number of activities
  – travel episodes
• replace rule set by behavioral model
  – interpretability
  – extrapolation

Thank you for your attention.