Sampling Urban Mobility through On-line Repositories of GPS Tracks

Michał PIóRKOWSKI
School of Computer and Communication Sciences
EPFL IC ISC LCA4
Motivation

• Mobility mining for:
  – Mobility modeling
  – Communication protocol design
  – Mobile system evaluation

• Mobility mining not only for networking:
  – Social science
  – Urban planning
  – Epidemiology
Large-Scale Mobility Data

• Types
  – *Contact-based* (Bluetooth)
  – *Track-based* (geo-location, AP/BS association)
  – *Usage-based* (phone calls, WiFi data traffic)

• *Where to get it from?*
  – Personal interviews [HsuMSHH05]
  – Telecoms, WLAN operators [BalazinskaC03], [GonzalesHB08]
  – Controlled experiments [EagleP06], [JangYZ08]
Large-Scale Mobility Data

• Problems
  – Scalability
  – Recruitment
  – User’s Privacy
  – Spatiotemporal Granularity

• Solution
  – Rely on self-contributed mobility traces available through on-line repositories of GPS tracks
Mobility Data

• Short-term mobility
  – Nokia Sports Tracker (NST) [http://sportstracker.nokia.com]
    • Since April, 2007
    • >2.5 million downloads (February 2009)
    • The best mobile Internet service (Mobile World Congress 2009)
  – Dataset:
    • 125’000 traces worldwide in kml files
    • 15 months: June 2007 – September 2008
    • Context: walking, running, cycling, driving, etc.
Mobility *Data*

- NST track data
- GPS errors unavoidable

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>user</td>
<td>nickname</td>
</tr>
<tr>
<td>type of activity</td>
<td>e.g. walking</td>
</tr>
<tr>
<td>start time</td>
<td>date hh:mm</td>
</tr>
<tr>
<td>duration</td>
<td>hh:mm:ss</td>
</tr>
<tr>
<td>traveled distance</td>
<td>km</td>
</tr>
<tr>
<td>average speed</td>
<td>km/h</td>
</tr>
<tr>
<td>maximum speed</td>
<td>km/h</td>
</tr>
<tr>
<td>average pace</td>
<td>1/km</td>
</tr>
<tr>
<td>maximum pace</td>
<td>1/km</td>
</tr>
<tr>
<td>the route</td>
<td>set of waypoints</td>
</tr>
</tbody>
</table>
Mobility *Data*

- Filtering spurious tracks

<table>
<thead>
<tr>
<th></th>
<th>Helsinki</th>
<th></th>
<th>Stockholm</th>
<th></th>
<th>London</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Raw</td>
<td>Filtered</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu_v$</td>
<td>$\hat{\sigma}_v^2$</td>
<td>$\hat{\mu}_v$</td>
<td>$\hat{\sigma}_v^2$</td>
<td></td>
</tr>
<tr>
<td>Walking</td>
<td>7.75</td>
<td>180.15</td>
<td>6.37</td>
<td>46.11</td>
<td></td>
</tr>
<tr>
<td>Running</td>
<td>10.83</td>
<td>551.24</td>
<td>9.85</td>
<td>4.10</td>
<td></td>
</tr>
<tr>
<td>Cycling</td>
<td>18.81</td>
<td>69.82</td>
<td>17.59</td>
<td>35.26</td>
<td></td>
</tr>
<tr>
<td>Walking</td>
<td>6.65</td>
<td>191.22</td>
<td>6.18</td>
<td>10.52</td>
<td></td>
</tr>
<tr>
<td>Running</td>
<td>11.19</td>
<td>45.59</td>
<td>10.35</td>
<td>11.89</td>
<td></td>
</tr>
<tr>
<td>Cycling</td>
<td>18.16</td>
<td>41.70</td>
<td>16.98</td>
<td>23.82</td>
<td></td>
</tr>
<tr>
<td>Walking</td>
<td>8.08</td>
<td>321.23</td>
<td>7.48</td>
<td>24.70</td>
<td></td>
</tr>
<tr>
<td>Running</td>
<td>11.05</td>
<td>11.70</td>
<td>11.07</td>
<td>6.33</td>
<td></td>
</tr>
<tr>
<td>Cycling</td>
<td>19.27</td>
<td>614.94</td>
<td>18.29</td>
<td>36.62</td>
<td></td>
</tr>
</tbody>
</table>

Mobility Data

- Helsinki: 731 users, 3757 traces
- Stockholm: 157 users, 1056 traces
- London: 504 users, 2488 traces
Mobility Data

- Is NST based data representative?
  - Utility cycling vs. recreational cycling
  - NST position art by Peter Rullmann

NST no longer Sports but Life Tracker

Large-Scale Mobility

Mobility Data

Applications

Mobility Mining
Mobility mining

- **Short-term vs. long-term mobility**
  - *Long-term*: locations of individuals recorded on regular basis independently on their activities
  - *Short-term*: occasional snapshots of long-term mobility

Do mobility patterns, inferred from short and long term mobility traces, share similar macroscopic characteristics?
Long-term mobility

- **Warsaw**
  [Sarafijanovic-DjukicPG06]
  - 825 cabs
  - 92 days
  - >517 km²

- **San Francisco**
  [PiorkowskiSDG09]
  - 536 cabs
  - 26 days
  - >600 km²
Spatial Heterogeneity
Spatial Heterogeneity
Connectivity
Mobile Partitioned Networks

- Clustered networks equally realistic as sparse and dense
- Cluster formation attributed to environmental features
- Heterogeneous Random Walk mobility model [PiorkowskiSDG09]

City-Specific Mobility Profiles

• Collection of places with specific mobility attributes:
  – Node density
  – Activity type
  – Travel time
  – …
Large-Scale Mobility

Mobility Data

Mobility Mining

Applications
Application: Location Privacy

• How to enhance user’s privacy in location-based services?
• Individuals use pseudonyms for identification
• Adversaries track user’s locations by *in-situ* pseudonym monitoring

• **Existing solution:** individuals should change pseudonyms in *Mix Zones* (regions obfuscating relation between entering and exiting individuals) [BeresfordS04]

Application: Location Privacy

• Improving Location Privacy by:
  – Increasing the number of individuals in Mix Zones
  – Placing Mix Zones at locations where individual mobility profiles are the same

• Optimal placement of Mix Zones [FreudigerSH09]
  – Adversary’s tracking success is decreased when Mix Zones are well placed

In reality, how do we know that a particular location is optimal for a Mix Zone?

Application: Location Privacy

Helsinki: Walking
Helsinki: Running
Helsinki: Cycling

Stockholm: Walking
Stockholm: Running
Stockholm: Cycling

London: Walking
London: Running
London: Cycling
Conclusions

• Novel method for gathering large scale mobility data
• Short-term vs. long-term mobility – similar macroscopic mobility characteristics
• City-specific mobility profiles inferred from collection of short-term mobility traces
• Improving privacy in mobile networks
Outlook

• More data: [flickr] [CrandallBHK09]
• Activity cycles
• Individual vs. collective mobility
• New context-aware services
• Malware propagation in mobile networks

References

• [BeresfordS04] A. Beresford, F. Stajano, Mix Zones: User Privacy in Location-aware Services, PerSec 2004
Conclusions

• Novel method for gathering large scale mobility data
• Short-term vs. long-term mobility – similar macroscopic mobility characteristics
• City-specific mobility profiles inferred from collection of short-term mobility traces
• Improving privacy in mobile networks

http://people.epfl.ch/michal.piorkowski