Deriving Tourist Mobility Patterns from Check-in Data

Linus W. Dietz, Daniel Herzog, Dr. Wolfgang Wörndl
Technical University of Munich
Department of Informatics
Chair of Connected Mobility
Los Angeles, California, USA
February 9, 2018
Introduction

Problem
Recommend composite trips of travel destinations

Motivation
Independent travel planning is complex, information is scattered, outdated, of uncertain quality
Introduction

Interactive Recommender Systems

• User modeling
• Recommendation algorithms
• UI/UX aspects
• Item characterization

Research questions

1. Which destinations to use in the travel knapsack (constrained by time & money)?
2. How long to stay at a specific region?
Solution

Deriving Tourist Mobility Patterns from Check-in Data

1. Which countries are frequently visited (together)
2. How long travelers typically stay at a specific region

Visualization of a trip © 2017 Google
Data

FOURSQUARE

Global check-in data set

• April 2012 to September 2013
• 266,909 users ⇒ 65,745 travelers
• 3,680,126 venues
• 77 countries
• 415 cities

Source: https://sites.google.com/site/yangdingqi/home/foursquare-dataset
Metrics

Trip duration

Check-in rate, check-in density

Transition time
Metrics

Trip duration

Check-in rate, check-in density

Transition time
Metrics

Trip duration

Check-in rate, check-in density

Transition time

Home $d_0 \rightarrow_3$

$\overset{\text{Block}}{\rightarrow_3} \overset{\text{Trip}}{d_3 \overset{\text{Brazil}}{\rightarrow_0} d_30} \rightarrow_1$

$\rightarrow_1 d_{31} Home$
Metrics

Trip duration

Check-in rate, check-in density

Transition time

$Home \ d_0 \rightarrow_3$

$\rightarrow_3 \ Block \ d_3 \ Brazil \ d_30 \rightarrow_1$

$\rightarrow_1 \ d_31 \ Home \ d_{100} \rightarrow_1$

$\rightarrow_1 \ Block \ d_{101} \ Netherlands \ d_{101} \rightarrow_0 \ Block \ d_{101} \ Germany \ d_{105} \rightarrow_1 \ Block \ d_{106} \ France \ d_{118} \rightarrow_3 \ Block \ d_{121} \ Germany \ d_{124} \rightarrow_0 \ Block \ d_{124} \ Netherlands \ d_{124} \rightarrow_6$

$\rightarrow_6 \ d_{130} \ Home$…
Data Processing

266,909 users

Filter out non-travelers

65,745 travelers

Filter out short trips (< 7 days)

23,218 travelers
34,892 trips

Filter out low check-in density users (< 0.2)

23,418 trips

Problem: high transition time (9.80 days)

Mean transition time: 3.39 days
## Results

### Country Co-occurrences

<table>
<thead>
<tr>
<th>n</th>
<th>Countries</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Malaysia, Singapore</td>
<td>269</td>
</tr>
<tr>
<td></td>
<td>Canada, USA</td>
<td>162</td>
</tr>
<tr>
<td></td>
<td>United Kingdom, France</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td>France, Italy</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>United Kingdom, USA</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>France, Italy, Spain</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>France, Belgium, Netherlands</td>
<td>38</td>
</tr>
<tr>
<td>3</td>
<td>United Kingdom, France, Italy</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>United Kingdom, France, Spain</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>Thailand, Malaysia, Singapore</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>United Kingdom, France, Italy, Spain</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>France, Belgium, Germany, Netherlands</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Austria, Czech Rep., Germany, Hungary</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>France, Belgium, Italy, Netherlands</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>France, Germany, Italy, Spain</td>
<td>10</td>
</tr>
</tbody>
</table>
Results  Durations of Stay

The graph shows the durations of stay for various countries. The x-axis represents the number of observations, and the y-axis represents the mean days. The countries are listed along the bottom edge of the graph, and the data points are marked with circles, with black dots indicating observations.

Linus W. Dietz (TUM)
Conclusions & Future Work

Check-in data $\Rightarrow$ tourist mobility patterns $\Rightarrow$ • durations of stay
  • destination co-occurrences

Metric-driven data mining approach $\Rightarrow$ Evaluate metrics with more data

Data models (destinations, trips) $\Rightarrow$ Fine-grained region model

Limited by data availability $\Rightarrow$ Combine several data sources

Refine heuristics

Determine traveler types
Conclusions & Future Work

Check-in data ⇒ tourist mobility patterns ⇒ • durations of stay
• destination co-occurrences

Metric-driven data mining approach

Data models (destinations, trips)

Limited by data availability

→ Evaluate metrics with more data
→ Fine-grained region model
→ Combine several data sources
Refine heuristics
Determine traveler types

Questions?

linus.dietz@tum.de
Deriving Tourist Mobility Patterns from Check-in Data

Linus W. Dietz, Daniel Herzog, Dr. Wolfgang Wörndl
Technical University of Munich
Department of Informatics
Chair of Connected Mobility
Los Angeles, California, USA
February 9, 2018
Data Model

Region tree of the planet

Earth
→ continents
   → countries
   → states
   → counties

GeoTree model

Continents
→ Continental sections
   → Countries
   → Regions
   → More regions
   → Cities
   → Districts

Wikitravel model
Data Model

Region tree of the planet

Earth → continents
  → countries
  → states
  → counties

GeoTree model

Continents → Continental sections
  → Countries
  → Regions
  → More regions
  → Cities
  → Districts

Wikitravel model

Current granularity on country level only!