Mobility Data Mining

Understanding Human Mobility
Mobility Profiles
Derived patterns and models

- Combination & refinement of basic patterns and models

- Individual Mobility Profile: routines consistently followed by a single moving object
User’s Mobility Profile

Given the user history as an ordered sequence of spatio-temporal points, we want to extract a set of routines in order to create the his/her mobility profile.

Where:

- A Routine is a typical local behavior of the user.
- A Mobility profile is the set of user’s routines
User’s Mobility Profile

Single trips of the user

Mobility profiles
Derived patterns and models: mobility profiles

User history
An ordered sequence of spatio-temporal points.

Trips construction
Cutting the user history when a stop is detected

Grouping
Performing a density based clustering equipped with a spatio temporal distance function

Pruning
Groups with a small Number of trips are Pruned

Profile extraction
The medoid of each group becomes user’s routines and the all set become the user’s mobility profile

Trasarti, Pinelli, Nanni, Giannotti. Mining mobility user profiles for car pooling. ACM SIGKDD 2011
What kind of distance?

- **Start + End**
  - Look for origin-destination pairs

- **Route similarity**
  - Look for recurrent paths followed

- **Temporal dimension**
  - Include time (of the day) to distinguish temporal regularity
What kind of representative?

- Classical average centroid cannot be applied
  - What is the centroid trajectory? Could make no sense

- Two practical solutions
  - Medoid: most central element of the cluster, e.g. minimized the sum of distances
  - Random: good enough if the clustering parameters are tight
Map Matching
Objective

How to transform this...
Gps raw trajectories
Avg sampling rate 90 seconds
Affected by GPS positioning error

...into this?
Sequences of road segments crossed
Objective

- Associate a sorted list of user positions to the road network on a digital map
- Two kinds of problems to solve
  - Map points to streets
  - Reconstruct path between points
Point mapping

- Determine which road segment a point belongs to
- Choose position within the segment

Input: GPS Track
Output: 'snapped' road
Objective

- Path reconstruction
  - Needed when gap between points is large
State of the art

- Map matching algorithms rely on shortest path between GPS points
Result

- Trajectory $\rightarrow$ sequence of road segments
Result

- Trajectory → sequence of road segments
Matching GPS data with shortest path leads to significant differences w.r.t. real GPS travel time between two points.
Alternative, Time-Aware approach

• Given a road network with travel times for each edge, find the path that best fits given total travel time
• Satisfy some basic constraints, e.g. no useless turnarounds

Projected length = 200 * cos(46°) = 155.95 m
Supposed time = \( \frac{\text{projected length}}{\text{line speed}} \) = 10.92s

Projected speed = \( \frac{\text{segment length}}{\text{supposed time}} \) = 18.31 m/s

Supposed length = \( \text{projected speed} \times \text{segment time} \) = 219.78 m
Segment time cost: \( |219.78 - 200| = 19.78 \)
Finding the travel time

\[ \Delta_{\text{Shortest Path}} = 129.78\% \]

\[ \Delta_{\text{Time Aware}} = 82.1\% \]
Effectiveness
Activity Recognition
Objective

- Infer the purpose and/or activity performed of trips and locations

- Two approaches
  - Consider what kind of activities can be performed in that area
  - Consider how the user behaves (when and how he reaches the area, etc.)
Recognition through Points-of-Interest

Given a dataset of GPS tracks of private vehicles, annotate trajectories with the most probable activities performed by the user.

Associates the list of possible POIs (with corresponding probabilities) visited by a user moving by car when he stops.

A mapping between POIs categories and Transportation Engineering activities is necessary.
The enrichment process

- **POI collection**: Collected in an automatic way, e.g. from Google Places.

- **Association POI – Activity**: Each POI is associated to a "activity". For example Restaurant → Eating/Food, Library → Education, etc.

- **Basic elements/characteristics**:
  - \( C(POI) = \{ \text{category, opening hour, location} \} \)
  - \( C(\text{Trajectory}) = \{ \text{stop duration, stop location, time of the day} \} \)
  - \( C(\text{User}) = \{ \text{max walking distance} \} \)

- **Computation of the probability to visit a POI/ to make an activity**: For each POI, the probability of "being visited" is a function of the POI, the trajectory and the user features.

- **Annotated trajectory**: The list of possible activities is then associated to a Stop based on the corresponding probability of visiting POIs
Lat; Lon
TimeStamp: Sun 10:55 – 12:05

Wd = 500 m

Bank: Mon – Fri [8:00 – 15:30]
Dentist: Mon – Sat [9:00 – 13:00] [15:30 – 18:00]
Church: Mon – Sat [18:00 – 19:00]
Sun [11:00 – 12:00]
Primary School: Mon – Sat [8:00 – 13:00]
Input & Output

- Stop: Lat; Lon
- TimeStamp: Sun 10:55 – 12:05

Wd = 500 m

Church ➤ Services [80%]
Bar ➤ Food [20%]

Pastry ➤ Food [100%]
Extraction of personal places from Twitter trajectories in Dublin area

The points of each trajectory taken separately were grouped into spatial clusters of maximal radius 150m. For groups with at least 5 points, convex hulls have been built and spatial buffers of small width (5m) around them. 1,461,582 points belong to the clusters (89% of 1,637,346); 24,935 personal places have been extracted.

Statistical distribution of the number of places per person

Examples of extracted places
Recognition of the message topics, generation of topical feature vectors, and summarization by the personal places

Topics have been assigned to 208,391 messages (14.3% of the 1,461,582 points belonging to the personal places)

1) Some places did not get topic summaries (about 20% of the places)
2) In many places the topics are very much mixed
3) The topics are not necessarily representative of the place type (e.g., topics near a supermarket: family, education, work, cafe, shopping, services, health care, friends, game, private event, food, sweets, coffee)
Activity Recognition

Individual Mobility Networks
How to synthesize Individual Mobility?

Mobility Data Mining methods automatically extract relevant episodes: locations and movements.
How to synthesize Individual Mobility?

- Basic approach: compute movement features of each trip
  - Length
  - Average speed or Duration
  - Bee-line length
  - Time of the day
  - ...

How to synthesize Individual Mobility?

- More advanced approach: consider overall mobility of the user
- First step: rank individual preferred locations
How to synthesize Individual Mobility?

Graph abstraction based on locations (nodes) and movements (edges)
How to synthesize Individual Mobility?

High level representation
Aggregation of sensitive data
Abstraction from real geography
From raw movement...
... to annotated data
The ABC classifier

1) Build from data an Individual Mobility Network (IMN)

2) Extract structural features from the IMN

3) Use a cascading classification with label propagation (ABC classifier)
Extracting the IMN

\[ \tau(0) = 25 \]
\[ \tau(9) = 8 \]
\[ \tau(11) = 3 \]

\[ \omega(0, 6) = 4 \]
\[ \omega(2, 1) = 2 \]
\[ \omega(1, 2) = 4 \]
Extracting the IMN

**Trip Features**
- Length
- Duration
- Time Interval
- Average Speed

**Network Features**
- Centrality
- Clustering coefficient
- Average path length
- Predictability
- Entropy
- Hubbiness
- Degree
- Betweenness
- Volume
- Edge weight
- Flow per location
Extracting the IMN

\[ \omega(2, 1) = 3 \]

\[ \omega(1, 2) = 2 \]

<table>
<thead>
<tr>
<th>from</th>
<th>to</th>
<th>weight</th>
<th>ccFrom</th>
<th>ccTo</th>
<th>duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0.22</td>
<td>0.12</td>
<td>10 min</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0.22</td>
<td>0.12</td>
<td>5  min</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0.12</td>
<td>0.22</td>
<td>4  min</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0.12</td>
<td>0.22</td>
<td>6  min</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0.12</td>
<td>0.22</td>
<td>4  min</td>
</tr>
</tbody>
</table>
ABC Classifier

• Principles:
  – The activities of a user should be predicted as a whole, not separately
  – Some activities are easy to classify
  – Other activities might benefit from contextual information obtained from previous predictions

• E.g.: a place frequently visited after work might be more likely to be leisure / shopping
ABC Classifier

- Reduce the multi-class problem into several binary problems
- The binary classifiers are learnt in cascade
- The classification results of each step are used as source for later classifications

Example

- Classifier 1: home vs all others
- Classifier 2: work vs all others
- Classifier 3: social activity vs all others
ABC Classifier

- Inspired by Nested Cascade Classification
ABC Classifier

- Inspired by Nested Cascade Classification
ABC Classifier

• After recognizing an activity (e.g. work), we use this information to enrich the features of the yet-unclassified trips

• E.g. add a feature describing whether the remaining trips are adjacent to the previous activity
  – Are there direct trips from work to the new place?
The ABC classifier
The ABC classifier
The ABC classifier
Experiments

GPS traces
6,953 trips
65 vehicles
## Experiments

<table>
<thead>
<tr>
<th>RF classifier</th>
<th>precision</th>
<th>recall</th>
<th>f1-score</th>
<th>activity</th>
<th>ABC classifier</th>
<th>precision</th>
<th>recall</th>
<th>f1-score</th>
<th>support</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.91</td>
<td>0.93</td>
<td>0.92</td>
<td>Going Home</td>
<td></td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>930</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Bring and get</td>
<td></td>
<td>1.00</td>
<td>0.68</td>
<td>0.81</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Education</td>
<td></td>
<td>1.00</td>
<td>0.25</td>
<td>0.40</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.14</td>
<td>0.14</td>
<td>Daily shopping</td>
<td></td>
<td>0.97</td>
<td>0.85</td>
<td>0.91</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.37</td>
<td>0.43</td>
<td>Working</td>
<td></td>
<td>0.93</td>
<td>0.96</td>
<td>0.94</td>
<td>258</td>
</tr>
<tr>
<td></td>
<td>0.34</td>
<td>0.6</td>
<td>0.43</td>
<td>Other</td>
<td></td>
<td>0.90</td>
<td>0.98</td>
<td>0.94</td>
<td>384</td>
</tr>
<tr>
<td></td>
<td>0.09</td>
<td>0.05</td>
<td>0.06</td>
<td>Shopping</td>
<td></td>
<td>0.87</td>
<td>0.85</td>
<td>0.86</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>0.13</td>
<td>0.09</td>
<td>0.11</td>
<td>Leisure</td>
<td></td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>0.24</td>
<td>0.11</td>
<td>0.15</td>
<td>Services</td>
<td></td>
<td>0.83</td>
<td>0.71</td>
<td>0.77</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.01</td>
<td>0.02</td>
<td>Touring</td>
<td></td>
<td>0.77</td>
<td>0.83</td>
<td>0.80</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>Food</td>
<td></td>
<td>0.76</td>
<td>0.49</td>
<td>0.60</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td>0.04</td>
<td>0.05</td>
<td>Social activities</td>
<td></td>
<td>0.65</td>
<td>0.69</td>
<td>0.67</td>
<td>49</td>
</tr>
<tr>
<td>avg / total</td>
<td>0.54</td>
<td>0.54</td>
<td>0.54</td>
<td></td>
<td></td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>1912</td>
</tr>
</tbody>
</table>
Experiments

- Is the order of activities in the learning relevant?

\[\text{mean} = 0.929\]
\[\text{std} = 0.016\]
Semantic Mobility Analytics

Temporal Analysis

- Pisa traffic
Semantic Mobility Analytics
Temporal Analysis

• Calci traffic
Semantic Mobility Analytics
Temporal Analysis

![Bar chart showing number of trips per hour across different categories]