Predicting Future Locations with Hidden Markov Models

Wesley Mathew, Ruben Raposo and Bruno Martins

Instituto Superior Técnico and INESC-ID
Technical University of Lisbon

4th International Workshop on Location-Based Social Networks
Predicting Future Locations with Hidden Markov Models

Wesley Mathew, Ruben Raposo and Bruno Martins

Table of contents

1 Introduction and Motivation
2 Related Work
3 The Proposed Location Prediction Method
4 Experimental Evaluation
5 Conclusions and Future Work
Analysis of human location histories

- Tasks like the prediction of the next location for a mobile user
- Mobile services with proactive context-based functions

Contributions from this research

- We describe and evaluate a method for predicting human mobility on the basis of Hidden Markov Models
- Experiments with a real-world location history dataset from the GeoLife project
Most human mobility trajectories show a high degree of temporal and spatial regularity, following simple and reproducible patterns (González et al., 2008).

Previous methods for the analysis of location histories
- **State-space models** - Markov models and related techniques
- **Data mining** - Sequential pattern mining techniques
- **Template matching** - Similarity search on sequential data
- **Other methods** - Combinations, probabilistic models, etc.

Examples of recently proposed methods
- Asahara et al. (2011) proposed a mixed Markov-chain model
- Methods competing in the Nokia 2012 Mobile Challenge
- Work presented at UbiComp 2012
Overview on the Proposed Method

Build models from a database of known trajectories

- Pre-process the human trajectory data
  - Each trajectory is a sequence of \(< \text{latitude}, \text{longitude}, \text{timestamp} >\)
  - Convert latitude and longitude into discrete region codes

- Group trajectories according to the temporal period in which they occurred
  - Three groups: weekdays daytime, weekdays nighttime, weekends

- Train a hidden Markov model for each group
  - Use Baum-Welch for estimating the parameters of the HMMs

Infer the most probable next-place for a new trajectory

- Pre-process the human trajectory data (i.e., the context)

- Inference over the corresponding HMM to estimate most probable next location
  - Use forward algorithm to compute the probability of possible sequences
  - Return the next place from the sequence with the highest probability
The Hierarchical Triangular Mesh

Recursive decomposition of a spherical Earth

- Start with 8 spherical triangles, 4 on each hemisphere
- Recursively sub-divide the triangles
  - Great circle arc segments connecting midpoints of each side
  - With $k$ steps, the number of regions is $n = 8 \times 4^k$

Learning of HMMs can be done more efficiently, using categorical distributions over regions instead of distributions over the real values for the geospatial coordinates.
Hidden Markov Models

- Well-known approach for the analysis of sequential data

Sequences (i.e., the human trajectories) are assumed to be generated by a Markov process with unobserved states

- Each state has a categorical distribution over possible regions
- Each state has a probability distribution encoding transitions other states
Learning the parameters of the model from data

- Given example sequences, find the best state transition and output probabilities
- Equivalent to deriving the maximum likelihood estimate for the parameters
- Local maximum likelihood derived efficiently using the Baum-Welch algorithm

Inferring the probability of a given sequence

- Requires computing a summation over all possible state sequences
- Problem can be handled efficiently using a dynamic programming procedure known as the forward algorithm

See the classic tutorial by Lawrence R. Rabiner (1989)
Experimental Protocol

Experiments with the GeoLife Dataset
- Data from 178 users collected from 2007-04 to 2011-10 with GPS devices
- Over 90% of trajectories with samples at every 15 seconds or every 10 meters
- Total distance of \( \approx 1.2 \) million kilometers and total duration of 48,000+ hours

Data pre-processing prior to experimentation
- Removed all duplicate consecutive locations from each trajectory
- Kept only the last 25 different locations that were visited
- Used 3465 different trajectories (i.e., one for each of the three clusters)

Experiments with different configurations of the proposed method
- Varying the number of states in the HMMs between 10, 15 and 20
- Varying the resolution of the triangular mesh between 18 and 22
- Different HMMs according to time period, versus a single HMM
## Results for Different Configurations

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Accuracy</th>
<th>Average Distance in Km</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>States</td>
<td>Res.</td>
</tr>
<tr>
<td>Multiple HMMs</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22</td>
</tr>
<tr>
<td>Single HMM</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22</td>
</tr>
</tbody>
</table>

- Best results with a single HMM, 15 states and a geospatial resolution of 18
- Distance errors are generally smaller for smaller values of geospatial resolution
  - Resolution of 18 corresponds to triangles of \( \approx 1280 \) squared meters, while resolution of 22 corresponds to triangles of \( \approx 5.002 \) squared meters
## Results for Different Temporal Periods

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Accuracy</th>
<th>Average Distance in Km</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P@1</td>
<td>P@5</td>
</tr>
<tr>
<td>Weekday 7AM to 7PM (Morning)</td>
<td>08.16</td>
<td>15.55</td>
</tr>
<tr>
<td>Weekday 7PM to 7AM (Evening)</td>
<td>04.90</td>
<td>16.47</td>
</tr>
<tr>
<td>Weekend</td>
<td>04.99</td>
<td>14.31</td>
</tr>
</tbody>
</table>
Distribution of the Distance Errors

![Graph showing the distribution of distance errors for Single HMMs and Multiple HMMs.](image.png)
Conclusions

- Described a method for predicting individual’s movements on the basis of Hidden Markov Models
  - A well-known approach for the analysis of sequential data
  - Accounting with location characteristics as unobservable parameters
  - Accounting with the effects of each mobile user’s previous actions

- Experiments conducted over the real-world location history dataset made available from the GeoLife project
  - Prediction accuracy of about 13.85% with the best performing method
  - Average distance of 143.506 Kilometers
  - Median distance of 4.957 Kilometers
Future Work

- Improve the training of HMMs through methods such as posterior regularization (Ganchev et al., 2010)
  - Incorporate indirect supervision through constraints

- More sophisticated clustering algorithms
  - Soft-clustering trajectories according to temporal period
  - Gaussian mixture models over the time-stamps

- More sophisticated models for sequential data
  - Higher-order HMMs
  - Non-parametric infinite-HMM model (Beal et al., 2002)
Thanks for your attention...

Questions?

bruno.g.martins@ist.utl.pt