Interdependence and Predictability of Human Mobility and Social Interactions

Antonio Lima
Joint work with Manlio De Domenico and Mirco Musolesi

Nokia MDC Workshop
Newcastle - June 18, 2012
Human Mobility. Can we predict it? We can, to a certain extent and at different geographic scales.
Research Questions

• Is it possible to improve the accuracy of the prediction by considering traces of multiple users?

• If yes, who should we select for improving the prediction of the movements of a given user?

• Can mobility correlation be considered as a cue for inferring social ties?
The Nokia MDC Dataset

• The complete dataset contains information from 152 smartphones (Nokia N95) for a year: address book, GPS, WLAN and Bluetooth traces, calls and SMS logs.

• We received data from 39 devices, 14 phone numbers were missing. We analysed a subset of the data related to 25 devices.
600 GPS (~60 hours) measurements (red) against forecast (black) for user 129

Linear Predictor

Error is of the order of 3 deg for lat-lng and 600 m for altitude.
Our Approach

- Multivariate nonlinear time series prediction.

- Extension to the case of multiple users. In particular pairs of users:
  - connected by a social link; and/or
  - with correlated mobility patterns.
The Mobility Model

\[ \dot{x}(t) = f(x, t) + \eta(t) \]

We consider the mobility model as a nonlinear dynamical system of a deterministic signal with a stochastic noise.

The position of a user is modeled with a 4-dim state vector.

We cannot analyse the d-dimensional phase space of the system directly.
The Mobility Model

We consider the mobility model as a nonlinear dynamical system of a deterministic signal with a stochastic noise.

The position of a user is modeled with a 4-dim state vector.

We cannot analyse the d-dimensional phase space of the system directly.

\[ \dot{x}(t) = f(x, t) + \eta(t) \]

User movements

\[ x_n = (h_n, \phi_n, \lambda_n, \xi_n) \]
The Mobility Model

We consider the mobility model as a nonlinear dynamical system of a deterministic signal with a stochastic noise.

The position of a user is modeled with a 4-dim state vector.

We cannot analyse the d-dimensional phase space of the system directly.
Takens’ Embedding Theorem

We can construct a space which preserves the dynamic properties of the system by using delayed measurements of the time-series. The theorem holds for noiseless time series of infinite length. We need a multivariate analysis to have a good precision on real-world time-limited noisy data.

\[ x_n \equiv (x_{n-(m-1)\tau}, \ldots, x_{n-\tau}, x_n) \]
Takens’ Embedding Theorem

We can construct a space which preserves the dynamic properties of the system by using delayed measurements of the time-series.

\[ \mathbf{x}_n \equiv (x_{n-(m-1)\tau}, \ldots, x_{n-\tau}, x_n) \]

\[ \mathbf{v}_n \equiv (y_{1,n-(m_1-1)\tau_1}, \ldots, y_{1,n}, y_{2,n-(m_2-1)\tau_2}, \ldots, y_{2,n}, \ldots, y_{M,n-(m_M-1)\tau_M}, \ldots, y_{M,n}) \]

The theorem holds for noiseless time series of infinite length. We need a multivariate analysis to have a good precision on real-world time-limited noisy data.
Takens’ Embedding Theorem

Embedding dimension = 8
\[ \mathbf{x}_n \equiv (x_n-(m-1)\tau, \ldots, x_n-\tau, x_n) \]
Delay time \( \sim 1000 \)

\[ \mathbf{v}_n \equiv (y_{1,n}-(m_1-1)\tau_1, \ldots, y_{1,n}, \]
\[ y_{2,n}-(m_2-1)\tau_2, \ldots, y_{2,n}, \]
\[ \ldots \]
\[ y_{M,n}-(m_M-1)\tau_M, \ldots, y_{M,n}) \]

We can construct a space which preserves the dynamic properties of the system by using delayed measurements of the time-series.

The theorem holds for noiseless time series of infinite length. We need a multivariate analysis to have a good precision on real-world time-limited noisy data.
Delay Embedding Reconstructions

Reconstruction for user 179

Reconstruction for a Brownian motion
Multivariate Nonlinear Prediction

\[ \hat{v}_{n+k} = \frac{1}{|U_n|} \sum_{v_j \in U_n} v_{j+k} \]

The prediction is performed considering the average over the states which are \( k \) steps ahead of the neighbours states.

In the reconstruction represented here for m=2, neighbours are inside the azure area.
Multivariate One-user Prediction

600 GPS (~60 hours) measurements (red) against forecast (black) for user 129

Much better, global prediction error of 0.19 deg for lat/lng and 219.43 m for the altitude.
Multivariate One-user Prediction

Much better, global prediction error of 0.19 deg for lat/long and 219.43 m for the altitude.

Let’s do even better!

600 GPS (~60 hours) measurements (red) against forecast (black) for user 129

Let’s do even better, global prediction error of 0.19 deg for lat/long and 219.43 m for the altitude.
Mobility Probability Density Function

PDF of positions of users who are friends (top) and who are not friends (bottom)
Mutual Information

\[ I(X, Y) = \sum_{x \in X} \sum_{y \in Y} P_{XY}(x, y) \log \frac{P_{XY}(x, y)}{P_X(x)P_Y(y)} \]

The mutual information quantifies how much information a stochastic variable can provide about another stochastic variable. It can be used as an estimator of the amount of correlation between them. If they are uncorrelated, it is null.

We use it to quantify how much the motion of a user can give us information about the motion of another.
Mutual Information, Contacts, Friendship

M.I. for people with at least one contact (left) and for people with no contacts at all (right).
Mutual Information, Contacts, Friendship

M.I. for people with at least one contact (left) and for people with no contacts at all (right).

Potential correlation with #contacts and friendship
The accuracy of the prediction improves by at least one order of magnitude (often two).

<table>
<thead>
<tr>
<th>Nodes</th>
<th>Social Link</th>
<th>Pos. Error [deg]</th>
<th>Alt. error [m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>026, 127</td>
<td>None</td>
<td>0.167</td>
<td>66.33</td>
</tr>
<tr>
<td>063, 123</td>
<td>Present</td>
<td>0.011</td>
<td>20.95</td>
</tr>
<tr>
<td>094, 009</td>
<td>Present</td>
<td>0.003</td>
<td>5.57</td>
</tr>
</tbody>
</table>
Take-away Messages

Human mobility traces are sometimes correlated.

Correlated traces improve forecasting accuracy.

Correlation can be a signal of social interaction.
Thanks!

Questions?

Antonio Lima

a.lima@cs.bham.ac.uk
http://cs.bham.ac.uk/axl162
@themiurgo

Thanks to flickr authors of CC-licenced images.
http://urli.st/M39