Time Intervals as a Behavioral Biometric

John (Vinnie) Monaco

Seidenberg School of CSIS, Pace University

November 11, 2015

http://vmonaco.com/dissertation
Outline

1 Introduction
   • Motivation
   • Background

2 Data
   • Description
   • Empirical patterns

3 Modeling
   • Model specification
   • Experimental results

4 Conclusions
“You are what *when* you eat”
Newell’s time scale of human action

<table>
<thead>
<tr>
<th>Scale (sec)</th>
<th>Time Units</th>
<th>System</th>
<th>World (theory)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^7$</td>
<td>Months</td>
<td>Task</td>
<td>SOCIAL BAND</td>
</tr>
<tr>
<td>$10^6$</td>
<td>Weeks</td>
<td>Task</td>
<td>SOCIAL BAND</td>
</tr>
<tr>
<td>$10^5$</td>
<td>Days</td>
<td>Task</td>
<td>SOCIAL BAND</td>
</tr>
<tr>
<td>$10^4$</td>
<td>Hours</td>
<td>Task</td>
<td>RATIONAL BAND</td>
</tr>
<tr>
<td>$10^3$</td>
<td>10 min</td>
<td>Task</td>
<td>RATIONAL BAND</td>
</tr>
<tr>
<td>$10^2$</td>
<td>Minutes</td>
<td>Task</td>
<td>RATIONAL BAND</td>
</tr>
<tr>
<td>$10^1$</td>
<td>10 sec</td>
<td>Unit task</td>
<td>COGNITIVE BAND</td>
</tr>
<tr>
<td>$10^0$</td>
<td>1 sec</td>
<td>Operations</td>
<td>COGNITIVE BAND</td>
</tr>
<tr>
<td>$10^{-1}$</td>
<td>100 ms</td>
<td>Deliberate act</td>
<td>COGNITIVE BAND</td>
</tr>
<tr>
<td>$10^{-2}$</td>
<td>10 ms</td>
<td>Neural circuit</td>
<td>BIOLOGICAL BAND</td>
</tr>
<tr>
<td>$10^{-3}$</td>
<td>1 ms</td>
<td>Neuron</td>
<td>BIOLOGICAL BAND</td>
</tr>
<tr>
<td>$10^{-4}$</td>
<td>100 µs</td>
<td>Organelle</td>
<td>BIOLOGICAL BAND</td>
</tr>
</tbody>
</table>
Behavioral biometrics.

The measure of human behavior for the purpose of identification or verification.
Timestamped events and time intervals.

- Timestamped events: keystrokes, touchscreen gestures, financial transactions, source code contributions...
- Given a series of events that occur at times $t_0, t_1, \ldots, t_N$

**Time interval between events**

$$\tau_n = t_n - t_{n-1}$$
Introduction
- Motivation
- Background

Data
- Description
- Empirical patterns

Modeling
- Model specification
- Experimental results

Conclusions
Why focus on timestamps?

- Timestamps are truly ubiquitous
- Timestamps are persistent
- Timestamps are resilient to encryption and masking
- Timestamps can generally be collected without cooperation
- Timestamps can be incorporated into domain-specific models
Problems.

**Identification**  Given a sequence of events, decide who they belong to (1 out of N)

**Verification**  Given a sequence of events with claimed responsibility, decide whether the claim is legitimate (binary classification)

**Prediction**  Given a sequence of events, predict the time of a future event
Bursts of activity in human behavior.

Random process (Poisson process, exponential inter-event times)

Bursty process (power-law inter-event times)

Barabasi, 2005
Time intervals of a random vs. bursty process.

Random process

Bursty process

Barabasi, 2005

John (Vinnie) Monaco

Time Intervals as a Behavioral Biometric
Psychology of human timing.

Implicit and explicit timing
Neurophysiology of human timing.

Praamstra, 2006

Wiener, 2011
Outline

1. Introduction
   - Motivation
   - Background

2. Data
   - Description
   - Empirical patterns

3. Modeling
   - Model specification
   - Experimental results

4. Conclusions
## Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Source</th>
<th>Size</th>
<th>Freq.(Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keystroke fixed-text</td>
<td>Monaco et al. (2013)</td>
<td>24k keystrokes, 60 users</td>
<td>4.4</td>
</tr>
<tr>
<td>Keystroke free-text</td>
<td>Villani et al. (2006)</td>
<td>251k keystrokes, 56 users</td>
<td>3.8</td>
</tr>
<tr>
<td>Mobile</td>
<td>Jain et al. (2014)</td>
<td>11k gestures, 52 users</td>
<td>3.1</td>
</tr>
<tr>
<td>Keypad</td>
<td>Bakelman et al. (2013)</td>
<td>6.6k keystrokes, 30 users</td>
<td>2.9</td>
</tr>
<tr>
<td>Bitcoin transactions</td>
<td>Reid et al. (2013)</td>
<td>239k transactions, 61 users</td>
<td>$2.8 \times 10^{-4}$</td>
</tr>
<tr>
<td>Linux kernel commits</td>
<td>Passos et al. (2014)</td>
<td>16k commits, 52 authors</td>
<td>$2.6 \times 10^{-6}$</td>
</tr>
<tr>
<td>White House visits</td>
<td>Hudson (2015)</td>
<td>2.7k visits, 18 people</td>
<td>$1.4 \times 10^{-6}$</td>
</tr>
<tr>
<td>Terrorist events</td>
<td>LaFree et al. (2007)</td>
<td>1.8k events, 10 groups</td>
<td>$2.8 \times 10^{-7}$</td>
</tr>
</tbody>
</table>
Keystroke.

Non-overlapping and overlapping keystrokes
Bitcoin transaction.

Transaction 1

1.0 BTC

Time $t_0$

0.75 BTC

0.25 BTC

Transaction 2

1.0 BTC

Time $t_1$

2.5 BTC

1.5 BTC
Terrorist activity.
Outline

1. Introduction
   - Motivation
   - Background

2. Data
   - Description
   - Empirical patterns

3. Modeling
   - Model specification
   - Experimental results

4. Conclusions
Heavy tails.

![Graphs showing heavy-tailed distributions for different datasets: Keystroke (fixed), Keystroke (free), Bitcoin, Kernel commits, White House visits, Terrorist activity.](image)
Preference for a log-normal.

Power law vs log-normal loglikelihood ratio tests

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Power law</th>
<th>Log-normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keystroke (free)</td>
<td>0.00 (0.00)</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>Keystroke (fixed)</td>
<td>0.00 (0.00)</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>Bitcoin</td>
<td>0.00 (0.00)</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>Kernel commits</td>
<td>0.75 (0.56)</td>
<td>0.25 (0.08)</td>
</tr>
<tr>
<td>White House visits</td>
<td>0.00 (0.00)</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>Terrorist activity</td>
<td>0.70 (0.20)</td>
<td>0.30 (0.00)</td>
</tr>
</tbody>
</table>
Time dependence.

- Keystroke (fixed)
- Keystroke (free)
- Bitcoin
- Kernel commits
- White House visits
- Terrorist activity

John (Vinnie) Monaco
Non-stationarity.
Temporal clustering.

**Description**

Empirical patterns

**Time Intervals as a Behavioral Biometric**
Outline

1. Introduction
   - Motivation
   - Background

2. Data
   - Description
   - Empirical patterns

3. Modeling
   - Model specification
   - Experimental results

4. Conclusions
Modeling approaches.

Windowed observations and event intensity

Time interval observations and event count

John (Vinnie) Monaco

Time Intervals as a Behavioral Biometric
Time interval distribution.

Log-normal

\[ f(\tau; \mu, \sigma) = \frac{1}{\tau \sigma \sqrt{2\pi}} \exp \left( -\frac{(\ln \tau - \mu)^2}{2\sigma^2} \right) \quad \tau > 0 \]
Transitioning between hidden states.

\[ z_t = 0 \]

\[ z_t = 1 \]

Active

Passive

\[ 1 - a_0 \]

\[ a_0 \]

\[ 1 - a_1 \]

\[ a_1 \]

John (Vinnie) Monaco

Time Intervals as a Behavioral Biometric
Hidden Markov model.

\[ z_1, z_2, z_{T-1}, z_T \]
\[ x_1, x_2, x_{T-1}, x_T \]

Time Intervals as a Behavioral Biometric
Partially-Observable Hidden Markov Model.

Partially-observable state

Hidden state

Observations

$z_1 z_2 z_{T-1} z_T$

$\omega_1 \omega_2 \omega_{T-1} \omega_T$

$z_1 z_2 z_{T-1} z_T$

$\omega_1 \omega_2 \omega_{T-1} \omega_T$

Observations

$0 1 2 T −1 T$

$x_1 x_2 x_{T-1} x_T$

Time Intervals as a Behavioral Biometric

John (Vinnie) Monaco
POHMM as an extension to the HMM.

- Introduces a dependency into the HMM to account for event types (e.g., key names).
- Can handle missing or incomplete observations by using the marginal distributions.
- Avoids overfitting through parameter mixing (or smoothing).
Consistency.

To be consistent the model must be:

- **Convergent**
  - Will our estimator always converge to a value?

- **Asymptotically unbiased**
  - Given a sample generated from a model with known parameters, can we recover the model parameters as the size of the sample increases?
Residuals.
Outline

1 Introduction
   • Motivation
   • Background

2 Data
   • Description
   • Empirical patterns

3 Modeling
   • Model specification
   • Experimental results

4 Conclusions
Evaluation criteria.

- Identification: rank-1 classification accuracy (ACC).
- Verification: equal error rate (EER), the point on the ROC curve where $P(\text{false accept}) = P(\text{false reject})$.
- Continuous verification: average maximum rejection time (AMRT), the average number of events before an impostor is detected without falsely rejecting the genuine user.
Evaluation procedure.

Folds

1
A  B  C
A

2
A  B  C
A

3
A  B  C
A

...

k
A  B  C
A

Reference

Query
Fitted model example.

Fixed-text

Free-text

John (Vinnie) Monaco
Time Intervals as a Behavioral Biometric
### Keystroke experimental results.

<table>
<thead>
<tr>
<th></th>
<th>Folds</th>
<th>Dichotomy</th>
<th>POHMM</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nursery rhymes</td>
<td>4</td>
<td>0.11 (0.04)</td>
<td><strong>0.00</strong> (0.01)</td>
<td>0.003</td>
</tr>
<tr>
<td>Keystroke (fixed)</td>
<td>4</td>
<td>0.13 (0.02)</td>
<td><strong>0.08</strong> (0.04)</td>
<td>0.041</td>
</tr>
<tr>
<td>Keystroke (free)</td>
<td>6</td>
<td><strong>0.02</strong> (0.01)</td>
<td>0.06 (0.01)</td>
<td>$8.9 \times 10^{-5}$</td>
</tr>
<tr>
<td>Keypad</td>
<td>20</td>
<td>0.11 (0.03)</td>
<td><strong>0.05</strong> (0.02)</td>
<td>$1.3 \times 10^{-8}$</td>
</tr>
<tr>
<td>Mobile (w/o sensors)</td>
<td>20</td>
<td>0.20 (0.03)</td>
<td><strong>0.10</strong> (0.02)</td>
<td>$2.7 \times 10^{-14}$</td>
</tr>
<tr>
<td>Mobile (w/ sensors)</td>
<td>20</td>
<td>0.01 (0.01)</td>
<td>0.01 (0.01)</td>
<td>0.500</td>
</tr>
</tbody>
</table>
Continuous verification.

**Fixed-text**

**Free-text**
Bitcoin experimental results.

- Hidden states are partially observable through the transaction direction (*incoming* or *outgoing*).
- 0.42 ACC
- 0.14 EER
- 139 AMRT
Linux kernel commit experimental results.

- Hidden states are partially observable through the commit intention (*bug fix or feature addition*).
- 0.17 ACC
- 0.36 EER
- 41 AMRT
White House visit experimental results.

- Hidden states are partially observable through the size of the group (*small* or *large*).
- 0.31 ACC
- 0.28 EER
- 19 AMRT
Terrorist activity experimental results.

- Hidden states are partially observable through the group intention.
- 0.15 ACC
- 0.45 EER
- 37 AMRT
What about anonymity?

- Timestamps can reveal your identity.
- Encryption, VPN, TOR, etc., cannot prevent that.
Alice and Bob want to be anonymous.
Masking strategy properties.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Finite</strong></td>
<td>The expected delay between the user and the arrival process should not grow unbounded.</td>
</tr>
<tr>
<td><strong>Anonymous</strong></td>
<td>The mix should make it difficult to identify the user.</td>
</tr>
<tr>
<td><strong>Unpredictable</strong></td>
<td>The mix should make it difficult to predict future behavior.</td>
</tr>
</tbody>
</table>
Proposed mixing strategies experimental results.

Masking capability increases as the tolerable lag increases.

**Keystroke (fixed)**

**Keystroke (free)**

**Bitcoin**

**Kernel commits**

**White House visits**

**Terrorist activity**

- Delay mix
- Interval mix
Conclusions.
Thank you