Using Event Data to Enhance Analytic Models

Jennifer Evans
Data Scientist
February 28, 2015
What is Event Data?

Records of an occurrence
Has a timestamp
Usually very granular

Types of Event Data?

Accidents  Upgrade Phone  Bird-Kicked Finch
Construction  Dropped Call  Bit Fatty
Rockslide  Make a Payment  Lost Almond
Avalanche  Add a Line  Found Almond
Road Closures  Check Contract End Date  Vet Visit
<table>
<thead>
<tr>
<th>TimeStamp</th>
<th>Identifier</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:10:11</td>
<td>Fatty</td>
<td>Bird Kicked Finch</td>
</tr>
<tr>
<td>1:10:11</td>
<td>Finch</td>
<td>Got Kicked</td>
</tr>
<tr>
<td>1:10:12</td>
<td>Finch</td>
<td>Bit Fatty</td>
</tr>
<tr>
<td>1:10:12</td>
<td>Fatty</td>
<td>Got Bit</td>
</tr>
<tr>
<td>2:10:10</td>
<td>Fatty</td>
<td>Found Almond</td>
</tr>
<tr>
<td>2:10:15</td>
<td>Finch</td>
<td>Stole Almond</td>
</tr>
<tr>
<td>2:10:15</td>
<td>Fatty</td>
<td>Lost Almond</td>
</tr>
<tr>
<td>2:10:17</td>
<td>Fatty</td>
<td>Bird Kicked Finch</td>
</tr>
<tr>
<td>2:10:17</td>
<td>Finch</td>
<td>Got Kicked</td>
</tr>
<tr>
<td>2:10:17</td>
<td>Finch</td>
<td>Lost Almond</td>
</tr>
<tr>
<td>2:10:20</td>
<td>Finch</td>
<td>Bit Fatty</td>
</tr>
<tr>
<td>2:10:20</td>
<td>Fatty</td>
<td>Got Bit</td>
</tr>
<tr>
<td>2:20:25</td>
<td>Tailess</td>
<td>Found Almond</td>
</tr>
<tr>
<td>5:30:15</td>
<td>Fatty</td>
<td>Vet Visit</td>
</tr>
</tbody>
</table>
## Sequence of Events by Bird

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Sequence</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatty</td>
<td>Bird Kicked</td>
<td>Got</td>
<td>Found</td>
<td>Lost</td>
<td>Bird</td>
<td>Got</td>
<td>Vet</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Finch</td>
<td>Bit</td>
<td>Almond</td>
<td>Almond</td>
<td>Kicked</td>
<td>Bit</td>
<td>Visit</td>
<td></td>
</tr>
<tr>
<td>Finch</td>
<td>Got Kicked</td>
<td>Bit</td>
<td>Stole</td>
<td>Got</td>
<td>Lost</td>
<td>Bit</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fatty</td>
<td>Fatty</td>
<td>Almond</td>
<td>Kicked</td>
<td>Almond</td>
<td>Fatty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tailess</td>
<td>Found Almond</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
- Fatty kicked Finch.
- Finch bit Fatty.
- Tailess found Almond.
### Analytic Data Set

<table>
<thead>
<tr>
<th>Bird</th>
<th>Bird Kicked Finch</th>
<th>Found Almond</th>
<th>Stole Almond</th>
<th>Bit Fatty</th>
<th>Got Bit</th>
<th>Got Kicked</th>
<th>bird-s</th>
<th>Event Sub-Frequencies</th>
<th>Event Frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatty</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>direct</td>
<td></td>
</tr>
<tr>
<td>Finch</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>indirect</td>
<td></td>
</tr>
<tr>
<td>Tailess</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

rpart(formula=Outcome~bird_kicked_finch+found_almond+stole_almond+bit_fatty+got_bit+got_kicked+bird_s+bird_kicked_finch_got_bit+got_kicked_bit_fatty,  
data=bird_day.df,  
method=“class”)
Variables You Might Want to Include

- Event Duration
- Duration Between Events
- Duration of Drive by Segments
- Time of Day by Segments
- Number of Events
- Number of Events in Subsequence
- Temperature
- Is it Snowing? (Y/N)
- Number of Accidents
Minds think with ideas, not information. No amount of data, bandwidth, or processing power can substitute for inspired thought.

— Clifford Stoll
Questions?

Jennifer.Evans@Clickfox.com