Data Mining and Anomaly Detection

David Belanger, PhD
Senior Research Fellow,
Stevens Institute of Technology
AT&T Labs Chief Scientist (retired)
Getting More Value from Data

- **Hundreds of Data Sources**
  - Volume
  - Velocity
  - Variety

- **Managing Risk:**
  - Security
  - Fraud
  - Compliance

- **Operations:**
  - Market Network Services
  - Customer Experience

- **Decision Support**

- **Process Control:**
  - Provisioning

- **Forecasting & Planning**

- **Data Analysis:**
  - Industry leading Information Mining Technology including RT

- **Information Visualization:**
  - The most effective ways to deliver Information & Alerts to decision makers?

- **Data In Flight:**
  - Efficient, Reliable, Secure Data Transport

- **Managed Services**

- **Vertical Services:**
  - Communications
  - Entertainment
  - Medical
  - Smart Grid
  - Hospitality

* Customer Specific Data is anonymized or aggregated.
Sources of Data - Communications

- **Network**
  - Circuit Switch: Signaling
  - IP: Netflow, Packets
  - Mobility, Wireless: RAN, RNC, Routers
- **Services**
  - Call Detail Records
  - Usage
- **Customers**
  - Billing
  - Customer Service
  - Surveys

- **Challenges – Alert/Alarm Systems**
  - Scale
  - Integrity
  - Security, Privacy
  - Efficiency
Sources of Data - Communications

1. Network Elements
2. Collectors
3. Network Recording
4. Near Real Time Analysis
5. Alerts, Alarms
6. Network Data Warehouses
7. Billing & Downstream Applications
Call Detail Record Data Analysis

Example Fields:
• Initiating Number
• Terminating Number
• Start Time and Date
• Call Duration
• Billing Number
• Phone Exchange
• Call Results – Busy, Failed, etc.
• Call Type – Voice, SMS, etc.

Example Uses of Network Recording Data:
• Billing
• Fraud
• Forecasting
• Media Event Voting
• Network Operations
• Customer Experience
• Non-syndromic
• Many more
Call Detail Record Data Analysis

Example Analytic Techniques - Fraud:

- Thresholds

<table>
<thead>
<tr>
<th>Call Duration &gt; X</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

- Signatures

<table>
<thead>
<tr>
<th>Day of Week</th>
<th>Region of World</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time of Day</td>
<td>% Countries Called</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Call Duration</th>
</tr>
</thead>
</table>

- Communities of Interest

![Diagram of communities of interest]
# Application of Data

<table>
<thead>
<tr>
<th>Input</th>
<th>Aggregate</th>
<th>Individual</th>
<th>Relational</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate</td>
<td>Wire Center, Zip, Cell Tower</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual</td>
<td>Thresholds</td>
<td>Signatures</td>
<td></td>
</tr>
<tr>
<td>Relational</td>
<td>Graph Measures</td>
<td>Col</td>
<td>Col</td>
</tr>
<tr>
<td>Crowd</td>
<td>Twitter</td>
<td>Mark the Spot</td>
<td>Social Network</td>
</tr>
</tbody>
</table>