Data-Driven Behavioral Analytics: Observations, Representations and Models

Meng Jiang (UIUC)
Peng Cui (Tsinghua)
Jiawei Han (UIUC)


Tutorial in CIKM 2016, October 24, Indianapolis, IN
What is Behavior?

- **Definition.** Interactions made by **individuals** in conjunction with **themselves** or their **environment**. (Wikipedia)
### Behavioral Analysis

#### Significance
What can we discover from behavioral data?

**Ex.** Given every phone call/message between military leaders, scientists, businesspersons, find...

<table>
<thead>
<tr>
<th>Observations</th>
<th>Who, what, where, when, why, how… (scientific view)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representations</td>
<td>Graph, network, matrix, tensor… (mathematical view)</td>
</tr>
<tr>
<td>Models</td>
<td>Prediction, recommendation, anomaly detection… (application view)</td>
</tr>
</tbody>
</table>
Why Behavioral Analysis Today?

*Today.* The human behaviors are broadly recorded in an unprecedented level. Insights of sciences and society?

**Physical World**

**Online Applications**
Basic Research Areas

- Six Disruptive Basic Research Areas
  - Engineered Materials (metamaterials and plasmonics)
  - Quantum Information and Control
  - Cognitive Neuroscience
  - Nanoscience and Nanoengineering
  - Synthetic Biology
  - Computational Modeling of Human and Social Behavior
VI. Computational Models of Human Behavior

A fundamental understanding and predictive capability of human behavior dynamics from individuals to societies.

- **Enabled capabilities**
  - Predictive models supporting strategic, operational, and tactical decision making and planning
  - Real time cultural situational awareness
  - Immersive training and mission rehearsal
  - Cross cultural coalition building

- **Key research challenges:**
  - Conflicting theories
  - Data management and fusion
  - Mathematical complexity
  - Validation of models

- **Measures of success**
  - Early success of simple models
  - Success of social network analysis
  - Prediction of crowd tipping points
Challenges in Behavioral Analysis

Content (preference)

Social context (influence)

Spatiotemporal context

Intention (suspiciousness)

Behavioral Analysis

<table>
<thead>
<tr>
<th>REWARDS</th>
<th># TICKETS GIVEN</th>
<th>CONSEQUENCES</th>
<th># TICKETS TAKEN AWAY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extra Math</td>
<td>+5</td>
<td>HITTING</td>
<td>-3</td>
</tr>
<tr>
<td>Getting kong (BULLYING)</td>
<td>+3</td>
<td>BULLYING</td>
<td>-4</td>
</tr>
<tr>
<td>Good Table Manners</td>
<td>+4</td>
<td>TEASING</td>
<td>-1</td>
</tr>
<tr>
<td>LOVE &amp; RESPECT</td>
<td>+5</td>
<td>LYING</td>
<td>-2</td>
</tr>
<tr>
<td>Checking the FIRST TIME</td>
<td>+3</td>
<td>THROWING A FIT</td>
<td>-3</td>
</tr>
<tr>
<td>Gain &amp; Quiet in STORE</td>
<td>+3</td>
<td>Ignoring Parents</td>
<td>-4</td>
</tr>
<tr>
<td>Extra Reading</td>
<td>+2</td>
<td>SPEAKING YELING</td>
<td>-1</td>
</tr>
<tr>
<td>CLEANING up after PLAYING</td>
<td>+2</td>
<td>BAD SPORT</td>
<td>-2</td>
</tr>
</tbody>
</table>
Methodology: Why Data-Driven?

- **Applications.** Recommender systems, fraud/spam detection.

- **Representation.** Behavior Network for interaction.
  - **Nodes:** users/authors, items (e.g., products, tweets, papers), etc.
  - **Links:** (interaction) following, purchasing, tweeting, publishing, etc.
  - **Node attributes:** user profiles, item properties/features, etc.
  - **Link attributes:** similarity, distance, weight, etc.
Data to Network to Knowledge

Information Network (entities, attributes, relationships)

Integration

Behavior Network

Rich unstructured text data
- tweets, news, msgs…
- product/restaurant review…
- publications (abstract/full text): PubMed, dblp, acm dl

Structured data
Outline: Data-Driven Behavioral Analytics

- Mining behavior networks with social and spatiotemporal contexts to support intelligent and trustworthy systems
  - Mining for behavior prediction and recommendation
  - Mining for suspicious behavior detection
Outline: Data-Driven Behavioral Analytics

- Mining behavior networks with social and spatiotemporal contexts to support intelligent and trustworthy systems
  - Mining for behavior prediction and recommendation
  - Mining for suspicious behavior detection
- Structuring behavioral content and integrating behavioral analysis with information networks
Acknowledgement
References


References


References


References


Y. Sun, R. Barber, M. Gupta, C. Aggarwar, and J. Han. “Co-author relationship prediction in heterogeneous bibliographic networks.” ASONAM, 2011.


J. Liu, J. Shang, C. Wang, X. Ren, and J. Han. “Mining quality phrases from massive text corpora.” SIGMOD, 2015.
References


References

Q. Li, Y. Li, J. Gao, B. Zhao, W. Fan, and J. Han. “Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation.” SIGMOD, 2014.


Thank you!

Data-Driven Behavioral Analytics: Observations, Representations and Models