Redirecting People with Complex Conditions to Effective Care

October 6, 2016

Center for Data Science & Public Policy

THE UNIVERSITY OF
CHICAGO
Johnson County Services

EMS

Mental Health Center

Jail, Court, Probation
County Services

127,000

575,000 people
“Frequent-flyer” explorations

EMS

Proportion of population

100%
80%
60%
40%
20%
0%

# of calls

- 9+
- 5-8
- 3-4
- 2
- 1

Jail

Days in jail

100%
80%
60%
40%
20%
0%

- 90+
- 10-89
- 2-9
- 1
- 0

Center for Data Science and Public Policy
University of Chicago
Prolonged interactions: Jail

Proportion of population

Days in jail
- 90+
- 10-89
- 2-9
- 1
- 0
Percent of population that entered jail in 2015
County Services

\[
\frac{575,000 \text{ people}}{4,430} \approx 130
\]
Sequence Analysis

Service Path Order:
E → E → E → E → E → M → E → J
Generalized Sequential Pattern

Specify time window size

Generate subsequence dataset

2 years, 1 years, 6 months, 3 months, 1 weeks, 3 days, etc.

Subsequence dataset
Demographics

- Age at earliest interaction with a public system
- Age group at last interaction with a public service

Counts of Interactions

- Number of bookings in last year
- Number of mental health entries in the last year
- Total number of bookings

Interaction Context

- Number of therapists seen
- Number of mental health services used
- Type of therapy
- Average bail amount

Timeline

- Standard deviation of time between public system interaction
- Had two bookings within a year
## Prioritized List: top 200 people

<table>
<thead>
<tr>
<th>Name</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matt Bauman</td>
<td>6.94</td>
</tr>
<tr>
<td>Eddie Lin</td>
<td>6.17</td>
</tr>
<tr>
<td>Kate Boxer</td>
<td>5.79</td>
</tr>
<tr>
<td>Erika Salomon</td>
<td>5.75</td>
</tr>
<tr>
<td>Lauren Haynes</td>
<td>5.02</td>
</tr>
<tr>
<td>Joe Walsh</td>
<td>4.72</td>
</tr>
<tr>
<td>Jen Helsby</td>
<td>4.49</td>
</tr>
<tr>
<td>Rayid Ghani</td>
<td>4.28</td>
</tr>
<tr>
<td>Steve Yoder</td>
<td>3.85</td>
</tr>
<tr>
<td>Chris Schnewies</td>
<td>3.64</td>
</tr>
<tr>
<td>Robert Sullivan</td>
<td>3.51</td>
</tr>
</tbody>
</table>
Johnson County : Outcomes
Prioritized List: top 200 people

52% (104) of the (top) 200 predicted individuals end up going to Jail in the next year

<table>
<thead>
<tr>
<th>Name</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matt Bauman</td>
<td>6.94</td>
</tr>
<tr>
<td>Eddie Lin</td>
<td>6.17</td>
</tr>
<tr>
<td>Kate Boxer</td>
<td>5.79</td>
</tr>
<tr>
<td>Erika Salomon</td>
<td>5.75</td>
</tr>
<tr>
<td>Lauren Haynes</td>
<td>5.02</td>
</tr>
<tr>
<td>Joe Walsh</td>
<td>4.72</td>
</tr>
<tr>
<td>Jen Helsby</td>
<td>4.49</td>
</tr>
<tr>
<td>Rayid Ghani</td>
<td>4.28</td>
</tr>
<tr>
<td>Steve Yoder</td>
<td>3.85</td>
</tr>
<tr>
<td>Chris Schnewies</td>
<td>3.64</td>
</tr>
<tr>
<td>Robert Sullivan</td>
<td>3.51</td>
</tr>
</tbody>
</table>
104 individuals

19 years total jail time

$250,000 absolute minimum cost

2 years since last mental health contact

<table>
<thead>
<tr>
<th>Name</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matt Bauman</td>
<td>6.94</td>
</tr>
<tr>
<td>Eddie Lin</td>
<td>6.17</td>
</tr>
<tr>
<td>Kate Boxer</td>
<td>5.79</td>
</tr>
<tr>
<td>Erika Salomon</td>
<td>5.75</td>
</tr>
<tr>
<td>Lauren Haynes</td>
<td>5.02</td>
</tr>
<tr>
<td>Joe Walsh</td>
<td>4.72</td>
</tr>
<tr>
<td>Jen Helsby</td>
<td>4.49</td>
</tr>
<tr>
<td>Rayid Ghani</td>
<td>4.28</td>
</tr>
<tr>
<td>Steve Yoder</td>
<td>3.85</td>
</tr>
<tr>
<td>Chris Schnewies</td>
<td>3.64</td>
</tr>
<tr>
<td>Robert Sullivan</td>
<td>3.51</td>
</tr>
</tbody>
</table>
Joco: How we got there

- Data Use agreements internally & Consolidation of data
- Co-responders
- Matching
- Anonymization
SLC: How we got there

- Data Use agreements internally & Consolidation of data
- Sharing data with UChicago
<table>
<thead>
<tr>
<th>Health Systems</th>
<th>Criminal Justice</th>
<th>Behavioral Health</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Paramedic transport logs</strong>&lt;br&gt;NOT: EMTs, ERs</td>
<td><strong>Jail bookings Court records Probation records</strong>&lt;br&gt;Not: arrests, dispatches</td>
<td><strong>County mental health case management</strong>&lt;br&gt;- Diagnoses&lt;br&gt;- Services&lt;br&gt;- Discharges</td>
</tr>
<tr>
<td><strong>Jail bookings Court records Probation records</strong>&lt;br&gt;Not: arrests, dispatches</td>
<td><strong>Mental health billings</strong>&lt;br&gt;- Services&lt;br&gt;- Dates</td>
<td><strong>Substance abuse treatment (TEDS)</strong></td>
</tr>
</tbody>
</table>

---

Center for Data Science and Public Policy
University of Chicago

dsapp.uchicago.edu
@datascifellows
Key data requirements

Venn diagram
- Requires linking individuals across multiple systems
  - Consistent combinations of identifying information can be used to match afterwards (more is better, must be normalized across silos):
    - Name fields
    - Date of birth
    - SSN
    - Gender
    - Race
    - Address

- Important to ensure data covers similar time periods
Key data requirements

Venn diagram

- Requires linking individuals across multiple systems
  - Consistent combinations of identifying information can be used to match afterwards (more is better, must be normalized across silos):
    - Name fields
    - Date of birth
    - SSN
    - Gender
    - Race
    - Address
- Important to ensure data covers similar time periods
- Additionally requires consistent collection and granularity of residence address to normalize by a total population
Key data requirements

Within-silo frequency
- One record per interaction
- Interaction type
- Consistent patient identifiers across interactions
  - Bonus points for linking repeated visits to the same ID at time of interaction
  - Otherwise, consistent combinations of identifying information can be used to match afterwards (more is better):
    - Name fields
    - Date of birth
    - SSN
    - Gender
    - Race
    - Address
Key data requirements

Across-silo frequencies

- One record per interaction
- Interaction type
- Date of interaction
- Duration of interaction

- Requires linking/matching from Venn diagram
Key data requirements

Sequence features
- Crucial to incorporate both start and stop dates of service/booking
- Note distinction between case management start/stop and interaction start/stop
- Documented reason of interaction end is valuable (good behavior, program drop out, administrative reasons, death)
Predictive analytics

- **ALL** available data from each silo
- We used over 300 indicators for each individual that were generated from the richness of the datasets of each silo:
  - Bail amounts
  - Type of therapy
  - Number of therapists
  - EMS primary impression

- Several (>3) years of history for all silos
- Existing baselines (thresholds) used to identify super-utilizers

**Note:**

- You can use the more robust “predictive” analysis to set more naive thresholds

<table>
<thead>
<tr>
<th>Name</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matt Bauman</td>
<td>6.94</td>
</tr>
<tr>
<td>Eddie Lin</td>
<td>6.17</td>
</tr>
<tr>
<td>Kate Boxer</td>
<td>5.79</td>
</tr>
<tr>
<td>Erika Salomon</td>
<td>5.75</td>
</tr>
<tr>
<td>Lauren Haynes</td>
<td>5.02</td>
</tr>
<tr>
<td>Joe Walsh</td>
<td>4.72</td>
</tr>
<tr>
<td>Jen Helsby</td>
<td>4.49</td>
</tr>
<tr>
<td>Rayid Ghani</td>
<td>4.28</td>
</tr>
<tr>
<td>Steve Yoder</td>
<td>3.85</td>
</tr>
<tr>
<td>Chris Schnewies</td>
<td>3.64</td>
</tr>
<tr>
<td>Robert Sullivan</td>
<td>3.51</td>
</tr>
</tbody>
</table>
Next Steps

• Asking for new jurisdictions
  – Data Use Agreements, Data & Data Transfer
  – Funding
• Tech Consortium Work
  – Open platform with the de-identified data to expand access to other researchers & analysts
Questions?

Contact:
• UChicago: Lauren Haynes (Lnhaynes@uchicago.edu)
• Johnson County:
  – Robert Sullivan (Robert.Sullivan@jocogov.org)
  – Chris Schneweis (Chris.Schneweis@jocogov.org)
  – Steve Yoder (Steve.Yoder@jocogov.org)
• Salt Lake County:
  – Fraser Nelson (Fnelson@slco.org)
Appendix
# Removing Identifying Information: hashing

One-way only

<table>
<thead>
<tr>
<th>Name</th>
<th>Hash</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lauren Haynes</td>
<td>e309c1ba03b22b72bc46cdf4200e0d19</td>
</tr>
<tr>
<td>Jane Doe</td>
<td>73c01bf88feb18695bd65e611ef1cf26</td>
</tr>
<tr>
<td>Matthew Bauman</td>
<td>18750ae79e3e94df96fdd4a354dbb2b0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phone</th>
<th>Hash</th>
</tr>
</thead>
<tbody>
<tr>
<td>655-82-8799</td>
<td>dfe1f020f8b457792a628692a607f94e</td>
</tr>
<tr>
<td>999-99-9999</td>
<td>007e3c1b311ec3848c878a420736c09f</td>
</tr>
<tr>
<td>000-00-0000</td>
<td>072f1bbdf1984fc0988be2d4b0c91803</td>
</tr>
</tbody>
</table>
Data Source

Aggregation


Prediction

Risk Score

Machine Learning

Random Forest
Gradient Boost
Logistic Regression
Decision Tree
Extra Tree
...

Feature1
Feature2
Feature3
...

Risk score for next year
Sequence Analysis

Service Path Order:
E → E → E → E → E → M → E → J
Generalized Sequential Pattern

Specify time window size

Generate subsequence dataset

2 years, 1 years, 6 months, 3 months, 1 weeks, 3 days, etc.

Subsequence dataset
Generalized Sequential Pattern

Specify time window size
Generate subsequence dataset
Break into subsets
Sequential pattern mining

Subsequence Dataset

GSP Algorithm

E→E→E
E→E→E→M→E→J
F→J→E→E
E→M→E→E→J
E→E→E→E→E→M
E→E→E→E→E→E
E→E→E→M→J
E→F→E→E
E→E→J→E→J→J
E→M→J
E→E→E→J
E→J→M→E→E→J
M→E→E→J
E→E→J→E→J
J→E→M→E→E→J
J→E→E→E→E
M→M
E→E→J→E
Generalized Sequential Pattern

E → E → E → E → E → M → E → J

?
Generalized Sequential Pattern

Specify time window size
Generate subsequence dataset
Break into subsets
Sequential pattern mining
Choose patterns to be features

Top-K Frequent Patterns in 6 months

Frequent Closed Sequence Mining without Candidate Maintenance, J. Wang, J. Han, and C. Li, IEEE Trans. on Knowledge and Data Engineering 19(8):1042-1056, IEEE Press, Piscataway, NJ, USA 2007
Predictive analytics
- **ALL** available data from each silo
- We used over 300 indicators for each individual that were generated from the richness of the datasets of each silo:
  - Bail amounts
  - Type of therapy
  - Number of therapists
  - EMS primary impression
- Several (>3) years of history for all silos
- Existing baselines (thresholds) used to identify super-utilizers
Anonymization, Hashing, Matching and Privacy

- The need for individual-level, linkable data
  - Anonymization Before Matching vs Anonymization after Matching
- Flow less sensitive information into more sensitive areas (Jail -> Mental Health)
- Hash anything that is otherwise public record (KDOC numbers) that would de-anonymize your data set
- Need a process owner for release of data (specific events that make the news can be de-anonymized)
Removing Identifying Information: Hashing

One-way only

<table>
<thead>
<tr>
<th>Name</th>
<th>Hash</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matt</td>
<td>0d2be2948013b41626bb3a43620ecb1c</td>
</tr>
<tr>
<td>matt</td>
<td>98287a567825c576a510b3edfbaa771c</td>
</tr>
<tr>
<td>Matthew</td>
<td>52f0f510aea2496759a4aa322099c21b</td>
</tr>
</tbody>
</table>