PDMP Patient Matching Challenges and Opportunities

Naresh Sundar Rajan, PhD, MS

Past: Senior Health Informaticist, Project Director, Prescription Drug Monitoring Program, Division of Occupation and Professional Licensing, State of Utah

Present: Interim – CTIO, & Senior Director, Interoperability & Informatics, Nebraska Health Information Initiative (NEHII)
Patient Matching – Perspective

• Universal problem – Many solutions
  • “No single solution to patient matching” – GAO, Jan. 2019 Report to Congress

• Challenge: Identifying the best rules/algorithm for your data

• Department of Health MPI (DOHMPI) (Probabilistic/Manually Curated)
  • Create a gold-standard MPI by linking different data sources across Utah, Vital Records, Cancer Registry, Controlled Substance Database, All-Payer Claims Database and etc.

• Utah Health Information Network (UHIN) – MPI (Referential)
  • A RESTful MPI Service to search patients across Utah’s population.
  • Authorized organizations can search using: Name, Gender, DOB, Address, Phone (Home, Work and Mobile) and SSN
  • Migrating to NextGate (third-party solution)
PDMP – Data Elements for Patient Matching

• Demographics (PATIENT TABLES)
  • First Name (required)
  • Last Name (required)
  • DOB (required)
  • Address (required)
  • City (required)
  • Zip-code (required)
  • Gender (optional)
  • Middle Name (optional)
Deep Probabilistic Patient Identity Resolution (DePPIR) – towards a data science approach

• Patient matching problem as a data science challenge
• Open-source (Python, PySpark, TensorFlow) tools
• Supervised Machine Learning based methods and annotated ASAP 4.2 version data model
• Hybrid approaches for blocking data to reduced pair-wise comparison by a significant number
• DePPIR – Open-sourced
DePPIR – Architecture

CSD PROD -> EXTRACT KEY DATA ELEMENTS -> ANNOTATE WITH PATIENT MATCHING DATA ELEMENTS

CSD STAGE -> DOMPHI/UHIN -> VALIDATION/EVALUATION

DePPIR BLACKBOX
- RANKING TOP 'k' MATCHES
- ADVANCED MACHINE LEARNING MODELS
- CUSTOM NICKNAME EDIT DISTANCE METHOD
- BASELINE MODELS
- PARAMETER TUNING
Current Stage – Evaluation

- **OBTAIN**: Gather data from relevant sources
- **SCRUB**: Clean data to formats that machine understands
- **EXPLORE**: Find significant patterns and trends using statistical methods
- **MODEL**: Construct models to predict and forecast
- **INTERPRET**: Put the results into good use

Originally by Hillary Mason and Chris Wiggins
Results

Demographics Data Elements
(PATIENT TABLES ASAP 4.2A)

First Name (required)*
Last Name (required)*
DOB (required)*
Address (required)*
City (required)
Zip-code (required)*
Gender (optional)*
Middle Name (optional)

Compared against Department of Health MPI (current gold-standard)

Sample Size: 13 Million Records
Models Used: Deep Neural Networks

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Precision Recall Gain</th>
<th>Area Under the Curve (AUC)</th>
<th>F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>DePPIR</td>
<td>99.68</td>
<td>99.84</td>
<td>95.84</td>
</tr>
</tbody>
</table>

Current model winners:

Incorrect addresses
Swapped middle names
Nicknames
Abbreviated names and addresses
Next Steps

• Include more features to gain F1 Score
  • Phone Number
  • SSN
• Explore feasibility of exposing DePPIR as a service (FHIR endpoints)
• Enhancing interstate PDMP Patient Matching by providing top-K matches for a given query
Points to Ponder

• Comparative effectiveness of Algorithms
• Understanding the quality of data
• Lack of transparency
• Ways to reduce human errors, create standardized data capture methods, and validity checks at the point of data ingestion
• Improve matching by including external sources such as biometrics (FastID), and Internet of Things (IoT)
Questions

• Naresh Sundar Rajan
• Email: nsrajan@nehii.org or nsr.informatics@gmail.com