Data-Driven Healthcare: Visual Analytics for Exploration and Prediction of Clinical Data

Adam Perer
IBM Research
Patient

Clinician

Electronic Medical Record Databases

**Thousands or Millions of Patients**

- 10+ Years of Data Per Patient
- Tens of Thousands of Features
  - Demographics
  - Diagnoses
  - Labs
  - Procedures
  - Claims
  - Unstructured Physician Notes
Patient Search and Analysis

Electronic Medical Record Databases

 Thousands or Millions of Patients
- 10+ Years of Data Per Patient
- Tens of Thousands of Features
  - Demographics
  - Diagnoses
  - Labs
  - Procedures
  - Claims
  - Unstructured Physician Notes
Patient

Clinician

Electronic Medical Record Databases

Expertise via Interaction

Thousands or Millions of Patients
• 10+ Years of Data Per Patient
• Tens of Thousands of Features
  • Demographics
  • Diagnoses
  • Labs
  • Procedures
  • Claims
  • Unstructured Physician Notes
• Visual tools for exploring clinical data support unearthing insights from clinical records

  • **CareFlow**

• Beyond exploration, clinical researchers often want predictions, too.

  • **Coquito, Prospector**
CareFlow
electronic medical records
electronic medical records

Patient #1

Antihypertensive
February 7, 2016
Patient #1

Antihypertensive
February 7, 2016

Beta Blockers
February 28, 2016

A

B

C

electronic medical records
electronic medical records

Patient #1

A

Antihypertensive
February 7, 2016

B

Beta Blockers
February 28, 2016

C

Diuretics
April 1, 2016
# electronic medical records

<table>
<thead>
<tr>
<th>Patient #1</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient #2</td>
<td>A</td>
<td>B</td>
<td>D</td>
</tr>
<tr>
<td>Patient #3</td>
<td>A</td>
<td>D</td>
<td>B</td>
</tr>
<tr>
<td>Patient #4</td>
<td>A</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>
electronic medical records

Patient #1

Patient #2

Patient #3

Patient #4

electronic medical records
heart failure

- Potentially fatal disease that affects 2% of adults in developed countries
- Difficult to manage
- No systematic clinical guidelines for treating Heart Failure
- Presence of co-morbidities affects treatment recommendations.
population

• Hundreds of Thousands of Patients diagnosed with Congestive Heart Failure
aggregation

- Start with target patient
- Find similar patients
  - Using our similarity analytics on relevant data
    - Features include medications, symptoms, and diagnoses, and lab tests
- Align all patients by disease diagnosis

What are the treatment pathways after diagnosis?
aggregation

Diagnosis
Date

[A]
[B]
[C]
[A,B]
[A,C]
[B,C]
[A,B,C]

Average outcome = 0.4
Average time = 10 days
Number of patients = 10
Hospitalized  Managed
Care Pathways of 300 similar patients
Optimal Care Pathway among 300 similar patients
other tools for clinical exploration

Videos and Papers at http://perer.org
clinical researchers:
clinical researchers:
clinical researchers:
clinical researchers:
clinical researchers:
the role of visualization in prediction
what can visualization do?

Cohort Definition

Model Interpretability

cocoito

prospector
defining cohorts

- Typically, defining cohorts is a slow process:
  - First, medical researchers define requirements.
  - Then, Technologists write SQL queries and deliver them to medical researchers.
  - But, often too many patients or too few patients, and the process must restart.
defining cohorts with coquito

drag and drop constraints with immediate feedback and hints for query refinement

supports using complex temporal logic

support for multiple queries side-by-side (for cases and controls)
coquito lessons

▪ Easy and interactive query formulation lets domain experts explore the data

▪ Visible intermediate results provide critical feedback

▪ Hints for query refinements are helpful in improving queries
typical predictive model report

Typically simply a list of top features and their weights

Why?
Difficult to summarize complex models

Issues
One cannot interpret how the values of each feature impact the prediction

One cannot interact with the model to test hypotheses
partial dependence

7/8 are predicted sick
partial dependence
partial dependence

demographic (age) - id: 100012 ix: 111 r: 0.868

age_at_enrollment (staticSum)

Cohort  Avg. Score
localized inspection

diabetes diagnoses: 7

bmi: 22

glucose level: 160

teeth: N

eyebrows: N

prediction: 0.95
localized inspection

- diabetes diagnoses: 7
- bmi: 20
- glucose level: 140
- teeth: N
- eyebrows: N
- prediction: 0.45
• predicting onset of diabetes for 4000 patients
• 4 month long term case study with 5 data scientists
• stories of visualization-driven insights in the paper
Patient: 5754  Truth: 1  Original: 0.71000  Current: 0.49000
take-aways

Clinical Data is complex and messy.

Exploratory visual analytics tools fill a much needed gap.

However, exploratory tools alone do not address their predictive desires.

There is a strong role for visualization in predictive tasks.

Adam Perer

[ papers and videos at http://perer.org ]