GAMAID: GREEDY CP TENSOR DECOMPOSITION FOR SUPERVISED EHR-BASED DISEASE TRAJECTORY DIFFERENTIATION

JETTE HENDERSON (UT-AUSTIN)
JOYCE C. HO (EMORY UNIVERSITY)
JOYDEEP GHOSH (UT-AUSTIN)
DIABETES & CHRONIC KIDNEY DISORDER

- Diabetes can cause diabetic nephropathy, a type of chronic kidney disease (CKD)
- 23% of diabetic patients suffer from CKD
  - Controlling glycemic levels are more challenging
  - Can result in complication of care

HOW DOES DIABETES LEAD TO KIDNEY DISEASE?

1. Diabetes leads to kidney disease in several ways.
   - At the onset of diabetes, blood flow into the kidneys increases, which may strain the glomeruli and lessen their ability to filter blood. Higher levels of blood glucose lead to buildup of extra material in the glomeruli, which increases the force of the blood moving through the kidneys and creates stress in the glomeruli.

2. This stress leads to gradual and progressive scarring of the glomeruli, eventually reducing the kidneys’ ability to filter blood properly.

3. Other factors including heredity, diet, lifestyle, and other medical conditions are also involved in the development of kidney disease.

OVERVIEW: GAMAID

- Exploratory, supervised method to separate diabetic patients into two groups based on risk of developing diabetic nephropathy.
- Model high-dimensional electronic health records (EHRs) using tensor to capture multi-way interaction (e.g., procedures used to treat diagnoses for a specific visit).
- Accumulate distinctive computational phenotypes that can differentiate patients with or without a disease.
EHR-BASED COMPUTATIONAL PHENOTYPES VIA TENSOR FACTORIZATION

EHR tensor to capture multi-way interaction of diagnoses and procedures
EHR-based computational phenotypes via tensor factorization

EHR tensor to capture multi-way interaction of diagnoses and procedures

Phenotypes 1

Diagnosis factor

Procedures factor

Phenotype importance

\( \mathbf{\Lambda} \)

\( \lambda_1 \)

\( \lambda_R \)

rank-one tensor

Hypertension Phenotype
(22% of patients)

- Bone/Joint/Muscle Infections/Necrosis
- Major Symptoms, Abnormalities
- Central Nervous System Infection
- Urinary Obstruction and Retention
- Surgical Procedures on the Female Genital System
- Microbiology Procedures
GAMAID: GREEDY CP TENSOR DECOMPOSITION

- Construct three tensors
  - One with data from both classes: $\mathcal{X}_{(01)}$
  - Two with data from each class: $\mathcal{X}_0, \mathcal{X}_1$
- Angular constraint to encourage diversity between discovered phenotypes
- Greedy algorithm to iteratively fit the best rank-one tensor
Step 1: Fit best rank-one approximation to tensor with both patients
Step 2: Fit best rank-one approximation to class 1 tensor that is distinct from the first rank-one approximation.
Step 2: Fit best rank-one approximation to class 1 tensor that is distinct from the first rank-one approximation.
Step 2: Fit best rank-one approximation to class 1 tensor that is distinct from the first rank-one approximation.
Step 3: Fit best rank-one approximation to class 0 tensor that is distinct from the other two rank-one approximation
Step R-1: Fit best rank-one approximation to class 1 tensor that is distinct from the other R-2 rank-one approximations
Step R: Fit best rank-one approximation to class 0 tensor that is distinct from the other R-1 rank-one approximations
DATA: CMS DE-SYNPUF

- CMS Linkable 2008-2010 Medicare Data Entrepreneurs’ Synthetic Public Use File (DE-SYNPUF)
- Focus on diabetic patients in 2009 and 2010
  - Class 1: 1,492 patients who did not have a CKD flag in 2009 but had one in 2010
  - Class 0: 1,625 patients who did not have a CKD flag in 2009 or 2010
- 3177 x 66 (diagnoses) x 198 (procedures)
GAMAID

DATA: CMS DE-SYNPUF

- CMS Linkable 2008-2010 Medicare Data Entrepreneurs’ Synthetic Public Use File (DE-SYNPUF)
- Focus on diabetic patients in 2009 and 2010
  - Class 1: 1,492 patients who did not have a CKD flag in 2009 but had one in 2010
  - Class 0: 1,625 patients who did not have a CKD flag in 2009 or 2010
- 3177 x 66 (diagnoses) x 198 (procedures)

diagnoses that appear in class 1 than class 0
## EXPERIMENTAL RESULTS: GAMAID

<table>
<thead>
<tr>
<th>Phenotype 1</th>
<th>% Class 1</th>
<th>% Class 0</th>
<th>% Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other gastrointestinal disorders</td>
<td>0.52</td>
<td>0.48</td>
<td>0.08</td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease and bronchiectasis</td>
<td>0.49</td>
<td>0.51</td>
<td>0.80</td>
</tr>
<tr>
<td>Fluid and electrolyte disorders</td>
<td>0.48</td>
<td>0.52</td>
<td>0.10</td>
</tr>
<tr>
<td>Abdominal pain</td>
<td>0.48</td>
<td>0.52</td>
<td>0.21</td>
</tr>
<tr>
<td>Chronic ulcer of skin</td>
<td>0.48</td>
<td>0.52</td>
<td>0.17</td>
</tr>
<tr>
<td>Other circulatory disease</td>
<td>0.54</td>
<td>0.46</td>
<td>0.09</td>
</tr>
<tr>
<td>Cardiac dysrhythmias</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Suture of skin and subcutaneous tissue</td>
<td>0.48</td>
<td>0.52</td>
<td>0.08</td>
</tr>
<tr>
<td>Routine chest X-ray</td>
<td>0.62</td>
<td>0.38</td>
<td>0.01</td>
</tr>
<tr>
<td>Other Laboratory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electrocardiogram</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonoperative urinary system measurements</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microscopic examination (bacterial smear, culture, toxicology)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other diagnostic procedures (interview, evaluation, consultation)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phenotype 6</th>
<th>% Class 1</th>
<th>% Class 0</th>
<th>% Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spondylosis; Intervertebral disc disorders; other back problems</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Physical therapy exercises, manipulation, and other procedures</td>
<td>0.48</td>
<td>0.52</td>
<td>0.08</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phenotype 9</th>
<th>% Class 1</th>
<th>% Class 0</th>
<th>% Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other gastrointestinal disorders</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other diagnostic radiology and related techniques</td>
<td>0.62</td>
<td>0.38</td>
<td>0.01</td>
</tr>
</tbody>
</table>
COMPARISON TO FISHER’S LDA

<table>
<thead>
<tr>
<th>Model</th>
<th>SVM F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA</td>
<td>0.4783</td>
</tr>
<tr>
<td>LDA + PCA</td>
<td>0.3914</td>
</tr>
<tr>
<td>GAMAID</td>
<td>0.5106</td>
</tr>
</tbody>
</table>
DISCUSSION + CONCLUSION

- Greedy CP decomposition with angular constraints
  - Potential to tease out phenotypes of diverging disease population
  - Exploratory method that produces easy to interpret results from high-dimensional data

- Future work
  - Understand tuning parameters (sparsity, fitting class 0 vs class 1, etc.)
  - Test framework on other sets of diseases
Q&A

- Contact information:
  - Jette Henderson (jette@ices.utexas.edu)
  - Joyce C. Ho (joyce.c.ho@emory.edu)
  - Joydeep Ghosh (jghosh@utexas.edu)