Cost-Sensitive Prediction: Applications in Healthcare

Daniel P. Robinson
Johns Hopkins University
Department of Applied Mathematics and Statistics

Collaborator: Suchi Saria (Computer Science)
Funding: IDIES Seed Funding Program

IDIES Annual Symposium
October 16, 2015
Cost-Sensitive Prediction: Applications in Healthcare

- The application: prediction of sepsis
- An exact cost-sensitive modeling formulation of the problem
- Convex relaxations
- Nonconvex relaxations and new optimization algorithms
Cost-Sensitive Prediction: Applications in Healthcare

- The application: prediction of sepsis
- An exact cost-sensitive modeling formulation of the problem
- Convex relaxations
- Nonconvex relaxations and new optimization algorithms
Cost-Sensitive Prediction: Applications in Healthcare

- The application: prediction of sepsis
- An exact cost-sensitive modeling formulation of the problem
- Convex relaxations
- Nonconvex relaxations and new optimization algorithms
Definition (sepsis)
Sepsis occurs when chemicals released into the bloodstream to fight an infection trigger inflammatory responses throughout the body.

Definition (septic shock)
Widespread infection causing organ failure & dangerously low blood pressure.

Facts:
- **200,000 – 3,000,000** cases of sepsis in the USA each year.
- Sepsis is estimated to cost American hospitals $20 billion each year.
- Most common and dangerous in older adults or those with weakened immune systems.
- Early treatment of sepsis with antibiotics and intravenous fluids improves chance of survival.
- **40%** of the patients diagnosed with sepsis do not survive.
- Early detection is key to improve survival rates.
The symptoms are not always straightforward or agreed upon.

Sepsis is a rare but serious condition that can look just like self-limiting infections such as flu, gastroenteritis or chest infections. See your GP immediately if you develop any one of the following:

- Slurred speech
- Extremely painful muscles
- Passing no urine (in a day)
- Severe breathlessness
- “I feel like I might die”
- Skin mottled or discoloured

email info@sepsistrust.org for more information
The application: prediction of sepsis

Cost structure graph

Costs at the activity level

Costs at the test level

Costs at the activity and test level

Layer 4 (Activities)
- insert-arterial-line
- none-1
- none-2

Layer 3 (Tests)
- arterial-bp
- non-invasive BP
- ekg-lead
- lactate
- blood-draw
- bmp

Layer 2 (Measurements)
- blood-pressure
- heart-rate
- lactate-level

Layer 1 (Features)
- shock-index
- 24hr-HR-mean
- lactate-level

Staff-time
Financial Cost
Wait time

$24
$35
$35
$24

20 min
20 min
30 min

AND
OR
Multiple Children
Research goals:

- prediction of septic shock
- model should capture the complex structure of the relevant costs
- evaluate convex/nonconvex relaxations of the exact formulation
- develop new algorithms to solve the resulting optimization problem
- evaluate performance on real healthcare data
- extensions to real-time setting, perhaps to aid in personalized healthcare
Cost-Sensitive Prediction: Applications in Healthcare

The application: prediction of sepsis

An exact cost-sensitive modeling formulation of the problem

Convex relaxations

Nonconvex relaxations and new optimization algorithms
Model prediction: Solve the problem

$$\min_{\beta} \mathcal{F}(\beta)$$

- $\beta \in \mathbb{R}^{\lvert f \rvert}$ is the weight vector
- For example, the logistic loss is

$$\mathcal{F}(\beta) = \sum_{n=1}^{N} \log \left( 1 + \exp \left( -y_n \beta^T f^n \right) \right)$$

- $f^n \in \mathbb{R}^{\lvert f \rvert}$ is the $n^{\text{th}}$ observed feature vector
- $y^n \in \{-1, 1\}$ the label of $f^n$
- $\lvert f \rvert$ the number of features
- $N$ the number of observations
To balance predictive accuracy and model complexity:

\[
\min_\beta \mathcal{F}(\beta) + \lambda C(\beta)
\]

- $\lambda$ is a weighting parameter
- Complexity/Cost function $C(\beta)$, e.g.,
  - $C(\beta) = \|\beta\|^2$
  - $C(\beta) = \|\beta\|_1$
  - $C(\beta) = \sum_{i \in G_j} \|\beta_{G_j}\|^2$

Key: none of the above cost regularizers accurately model the relevant costs
- first nonzero feature to enter is most predictive
- the most predictive may be the most expensive
- no control over the relevant costs
- a regularizer based on the complicated cost structure is needed!
A cost-sensitive regularizer

- $A$ nodes at activity layer
- $T$ nodes at test layer
- $M$ nodes at measurement layer
- $F$ nodes at feature layer

Reduce 4 layer boolean circuit (cost graph) to a 3-layer boolean circuit:

- layer-1 contains the nodes $F$
- layer-2 contains the nodes $Z := \{ f_{i,p} : 1 \leq i \leq n_f \text{ and } 1 \leq p \leq w_i \}$
- layer-3 contain the nodes $A$

- only or gate functions are used in layer-1
- only and gate functions are used in layer-2
A cost-sensitive regularizer

We now have that the cost for the $l$th care-giver activity is given by

$$C^a_l \ I \left( \bigvee_{(i,p) \in S^a_l} \beta_{i,p} \right)$$

where the index set $S^a_l$ is defined as

$$S^a_l := \{(i,p) : \text{the output of } g_{f_{i,p}}(\cdot) \text{ depends on } a_l\}$$

Comments:

- The definition of our regularizer follows only from the knowledge of the 3-layer (reduced) boolean circuit.
- This procedure generalizes to any cost graph that can be represented as a finite layer boolean circuit.
We then have the following optimization problem based on the exact regularizer:

$$\min_{\beta} \mathcal{F}(\beta) + \lambda C_{\text{exact}}(\beta)$$

- $\lambda$ is a weighting parameter
- Complexity/Cost function $C(\beta)$:

$$C_{\text{exact}}(\beta) := \sum_{l=1}^{|\mathcal{A}|} C_{l}^a I \left( \bigvee_{(i,p) \in \mathcal{S}_{l}^a} \beta_{i,p} \right) + \sum_{l=1}^{|\mathcal{T}|} C_{l}^t I \left( \bigvee_{(i,p) \in \mathcal{S}_{l}^t} \beta_{i,p} \right)$$

Comments:

- This cost-regularizer is exact, but not tractable for large-scale problems.
- Convex and nonconvex relaxations are possible.
Outline

1 Cost-Sensitive Prediction: Applications in Healthcare
   - The application: prediction of sepsis
   - An exact cost-sensitive modeling formulation of the problem
   - Convex relaxations
   - Nonconvex relaxations and new optimization algorithms
One could consider **convex relaxations**:

\[
\min_{\beta} \mathcal{F}(\beta) + \lambda C(\beta)
\]

where

\[
C(\beta) := \sum_{l=1}^{A} C_{l}^{a} \| \bigvee_{(i,p) \in S_{l}^{a}} \beta_{i,p} \|_{\infty} + \sum_{l=1}^{T} C_{l}^{t} \| \bigvee_{(i,p) \in S_{l}^{t}} \beta_{i,p} \|_{\infty}
\]

**Comments:**

- Sum of (overlapping) group regularizers.
- Can use software such as SPAMS to solve this problem.
- Tried this approach on the following data . . .
Setup:

- Used MIMIC-II, a large publicly available dataset containing electronic health records from patients admitted to the ICU at the Beth Israel Deaconess Medical Center over a period of seven years.
- Processed the data from all adults (age > 15) with at least one measurement of blood urea nitrogen, hematocrit, and heart rate.
- This yielded data from 16,232 patients.
- Septic shock onset was identified using the 2012 Surviving Sepsis Campaign definitions, which resulted in 2,291 patients.
- Patients with severe sepsis who never developed septic shock but received treatment, were removed since their outcome was confounded.
- Split individuals into training (75%) and test (25%) sets.
- Since the dataset is highly unbalanced, during training, we subsample the negative pairs to generate a balanced training set.
Table: Various costs for different models obtained by using our structured regularizer.

<table>
<thead>
<tr>
<th>Models</th>
<th>$\mathcal{M}_1$</th>
<th>$\mathcal{M}_2$</th>
<th>$\mathcal{M}_3$</th>
<th>$\mathcal{M}_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity at 0.85</td>
<td>0.61</td>
<td>0.66</td>
<td>0.65</td>
<td>0.72</td>
</tr>
<tr>
<td>specificity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>82.79 ± 0.55</td>
<td>84.45 ± 0.64</td>
<td>84.75 ± 0.55</td>
<td>87.21 ± 0.46</td>
</tr>
<tr>
<td>Financial Cost</td>
<td>$0$</td>
<td>$0$</td>
<td>$72$</td>
<td>$170$</td>
</tr>
<tr>
<td>Care-giver Time</td>
<td>0 minutes</td>
<td>10 minutes</td>
<td>0 minutes</td>
<td>30 minutes</td>
</tr>
<tr>
<td>Result Time</td>
<td>0 minutes</td>
<td>10 minutes</td>
<td>50 minutes</td>
<td>50 minutes</td>
</tr>
<tr>
<td>Tests Required</td>
<td>routine</td>
<td>routine, urine</td>
<td>abg, routine</td>
<td>abg, cbc, cmp, hct, hemoglobin, routine, urine</td>
</tr>
<tr>
<td>Activities Required</td>
<td>none</td>
<td>urine</td>
<td>arterial stick</td>
<td>arterial stick, blood draw, urine</td>
</tr>
</tbody>
</table>

Comments:

- Diverse models are easily obtained using this cost-sensitive regularizer.
- Easily can adjust model preferences.
Cost-Sensitive Prediction: Applications in Healthcare

1. The application: prediction of sepsis
2. An exact cost-sensitive modeling formulation of the problem
3. Convex relaxations
4. Nonconvex relaxations and new optimization algorithms
Solve the problem

$$\min_{\beta} \mathcal{F}(\beta) + \lambda C(\beta)$$

for choices such as

$$C(\beta) := \sum_{l=1}^{|A|} C_l^a \sqrt{\sum_{(i,p) \in S_l^a} |\beta_{i,p}|} + \sum_{l=1}^{|T|} C_l^t \sqrt{\sum_{(i,p) \in S_l^t} |\beta_{i,p}|}$$

Other choices are possible:

- Sigmoid functions.
- Min functions such as $\min\{0, \sum_{(i,p) \in S_l} |\tilde{\beta}_{i,p}|\}$. 
The general problem formulation

\[
\min_{\beta \in \mathbb{R}^{|f|}} \ F(\beta) + \sum_{j=1}^{|G|} g_j \left( \sum_{i \in U_j} |\beta_i| \right)
\]

- \( f : \mathbb{R}^{|f|} \to \mathbb{R} \) is convex
- \( g_i : \mathbb{R} \to \mathbb{R} \) is concave and increasing on \([0, \infty)\)

Comments:
- Difference of convex functions.
- Known algorithms when the gradient of \( F \) is cheap to compute.
- Methods become inefficient in the big data setting.
- We are developing methods that . . .
  - only use stochastic or batch gradients of \( F \) (scalable)
  - utilize the difference of convex functions structure
  - utilize projections onto cones
  - inexact subproblem solvers (efficiency)
  - have convergence guarantees
  - details . . . hopefully next year!
Thank You