Active Learning for Developing Personalized Treatment

Kun Deng\(^1\) Joelle Pineau\(^2\) Susan Murphy\(^1\)

\(^1\)Department of Statistics
University of Michigan

\(^2\)Department of Computer Science
McGill University

July 16, 2011
Outline

1. Motivation
   - Basic Problem

2. Methods and Algorithms
   - Optimization Criterion 2

3. Results and Discussion
   - Experimental Results for Criterion 2
   - Discussion
Patients are categorized into subpopulations $c_1 \sim c_4$ based on biomarkers. Two treatment actions $a_1$ and $a_2$

An individualized treatment rule (ITR) looks like:

$$d(c_i) = \begin{cases} a_1 & \text{if } \hat{\mu}_{i1} - \hat{\mu}_{i2} \geq 0 \\ a_2 & \text{if } \hat{\mu}_{i1} - \hat{\mu}_{i2} < 0 \end{cases} \quad \forall i \in \{1, 2, 3, 4\}$$

$\hat{\mu}_i$ are the sample mean responses for subpopulation $c_i$

An uncertainty measure in the estimated treatment effect: $\text{Var}[\hat{\mu}_{i1} - \hat{\mu}_{i2}] = \text{Var}[\hat{\mu}_{i1}] + \text{Var}[\hat{\mu}_{i2}]$ for each $i$.

A confidence measure in the correctness of the policy: $\text{Pr}[\hat{\mu}_{i1} > \hat{\mu}_{i2}]$, if, say, treatment 1 is the best for all subpopulations.
Introduction

- **Personalized Medicine/Treatment**
  - treat each patient based on his characteristics: patients with different gene biomarker or clinical biomarkers often show differential responses to the same treatment.
  - adapt treatment over time (not covered in this talk)

- Our Goal: collect reliable evidence for medical decision making
  - construct decision rules that are tailored to individual heterogeneity
  - quantify and optimize the quality of these decision rules in terms of their uncertainty, confidence of correctness etc.
  - make better use of limited clinical trial resources: number of people recruited
Current Practice and Discussion

Recruit from the entire population as patients arrive: patients in the trial roughly reflect their natural composition. A post subgroup analysis is used to derive treatment assignment for subpopulations.

The results lack power, are difficult to reproduce, because the trial is not powered to detect treatment differences in subpopulations.

Question: how to intelligently recruit patients from subpopulations in order to construct a more-balanced treatment policy.
**Our Approach**

- A minimax bandit model that intelligently recruits patient from different subpopulations and assigns them to different treatments
- Two performance criteria in terms of the quality of the treatment policy:
  - (Minimize) the largest variance of the estimated treatment effects among the different subpopulations
  - (Minimize) the probability of selecting suboptimal treatments across the different subpopulations
- Other performance criteria are possible too.
Assumptions

- Active treatment period of a patient is short compared to the pace of patient recruitments (i.e. the entire trial)
- Patient treatment and monitoring are very costly
- The budget for a clinical trial is specified a priori, say $N$ subjects maximally
A MiniMax Bandit Problem

- There are $C$ bandits (corresponding to the $C$ subpopulations), each equipped with $K$ arms.
- At each time point, we are only allowed to pick one bandit. For that bandit, we need to further decide an arm to pull.
- Mean $\mu_{ij}$ (corresponding to the primary outcome of action $(i,j)$) and variance $\sigma_{ij}^2$.
- Define some kind of loss, based on our goal of creating good ITRs, we want to control the maximum loss for all subpopulations.
- Focus on the loss regarding the confidence of the correctness of the ITRs.
Criterion 2: controlling maximal error probability of selection

Some Definitions

- Assume there is a single best treatment for each subpopulation $j^*_i$
- Define loss for a bandit (subpopulation) $i$

$$L^n_i = \Pr[\max_{j \neq j^*} \hat{\mu}_{ij} \geq \hat{\mu}_{ij^*}]$$

- The overall loss of an active learning policy $\pi$:
  $$L^n(\pi) = \max_{1 \leq i \leq C} L^n_i$$
- Aims to control the maximal error of incorrectly selecting a suboptimal treatment for patient of any subpopulations.
Motivation
Methods and Algorithms
Results and Discussion

Cont’d

- $L_i$ has a closed form, but not convex in $n_i$, neither is $\max_i L_i$.
- First, consider a surrogate oracle algorithm that knows mean/variance

\[
\Pr[\max_{j \neq j^*} \hat{\mu}_{ij} \geq \hat{\mu}_{ij^*}] \leq \sum_{j \neq j^*} \Pr[\hat{\mu}_{ij} \geq \hat{\mu}_{ij^*}] \leq \sum_{j \neq j^*} \frac{\nabla (\hat{\mu}_{ij} - \hat{\mu}_{ij^*})}{(\mu_{ij} - \mu_{ij^*})^2},
\]

surrogate: minimize

\[
\max_i \sum_{j \neq j^*} \frac{\sigma_{ij}^2}{n_{ij}} + \frac{\sigma_{ij^*}^2}{n_{ij^*}} (\mu_{ij} - \mu_{ij^*})^2
\]

s.t.

\[
\sum n_{ij} = N.
\]
The optimal surrogate oracle allocation is:

\[ n_{ij}^* = \frac{v_{ij} \sum_j v_{ij}}{\sum_i (\sum_j v_{ij})^2} N, \]

where

\[
\begin{align*}
\hat{v}_{ij}^2 &= \frac{1}{(\mu_{ij}^* - \mu_{ij})^2} \sigma_{ij}^2, & j \neq j^* \\
\hat{v}_{ij^*}^2 &= \sum_j \neq j^* \frac{1}{(\mu_{ij^*} - \mu_{ij})^2} \sigma_{ij^*}^2, & j = j^*.
\end{align*}
\]

We use \( \hat{\sigma}_{ij} \) and \( \hat{\mu}_{ij} \) to derive an active learning policy MINIMAXPICS, the next bandit/arm pulled is drawn according to:

\[
\left\{ \frac{\hat{v}_{ij} \sum_j \hat{v}_{ij}}{\sum_i (\sum_j \hat{v}_{ij})^2} ; i \in \{1, \ldots, C\}, j \in \{1, \ldots, K\} \right\}.
\]
We evaluate two variants against random sampling/assignment (AARandom):

MINMAXPICS(SEQ): \{\hat{v}_{ij} \sum_j \hat{v}_{ij} , 1 \leq i \leq C, 1 \leq j \leq K}\}

MINMAXPICS(GRP) selects the next subpopulation: \{\left(\sum_j \hat{v}_{ij}\right)^2 , 1 \leq i \leq C\} and randomly assigns one patient to each subpopulation. Why?

**Table:** Datasets for the MINMAXPICS comparison

<table>
<thead>
<tr>
<th>DS</th>
<th>subpop.</th>
<th>dist.</th>
<th>means</th>
<th>variances</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>subpop. treatments</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ds21</td>
<td>4/3</td>
<td>.25</td>
<td>20 10 10</td>
<td>50 50 50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.25</td>
<td>20 10 10</td>
<td>50 50 50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.25</td>
<td>20 10 10</td>
<td>50 50 50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.25</td>
<td>19 15 15</td>
<td>50 50 50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.25</td>
<td>20 10 10</td>
<td>50 50 50</td>
</tr>
<tr>
<td>Ds22</td>
<td>4/3</td>
<td>.25</td>
<td>20 15 15</td>
<td>50 50 50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.25</td>
<td>20 10 10</td>
<td>50 50 50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.25</td>
<td>20 10 10</td>
<td>50 50 50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.25</td>
<td>20 10 10</td>
<td>50 50 50</td>
</tr>
<tr>
<td>Ds23</td>
<td>5/3</td>
<td>.25</td>
<td>20 15 15</td>
<td>50 50 50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.25</td>
<td>20 10 10</td>
<td>50 50 50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.25</td>
<td>20 10 10</td>
<td>50 50 50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.25</td>
<td>20 10 10</td>
<td>50 50 50</td>
</tr>
<tr>
<td>Ds24</td>
<td>8/3</td>
<td>.125</td>
<td>20 15 15</td>
<td>50 50 50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.125</td>
<td>20 10 10</td>
<td>50 50 50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.125</td>
<td>20 10 10</td>
<td>50 50 50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.125</td>
<td>20 10 10</td>
<td>50 50 50</td>
</tr>
<tr>
<td>Ds2-CBASP</td>
<td>3/2</td>
<td>1.5</td>
<td>10.9 16.2</td>
<td>99.3 79.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2/5</td>
<td>9.3 19.4</td>
<td>110.7 55.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2/5</td>
<td>12.9 15.8</td>
<td>103.5 78.6</td>
</tr>
</tbody>
</table>
Experimental Results for Criterion 2

DS21

AARandom
MINMAXPICS (SEQ)
MINMAXPICS (GRP)
Oracle Allocation (surrogate)

DS22

AARandom
MINMAXPICS (SEQ)
MINMAXPICS (GRP)
Oracle Allocation (surrogate)

DS23

AARandom
MINMAXPICS (SEQ)
MINMAXPICS (GRP)
Oracle Allocation (surrogate)
## Table: Datasets for the MINMAXPICS comparison

<table>
<thead>
<tr>
<th>DS</th>
<th>subpop./dist.</th>
<th>means</th>
<th>variances</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS21</td>
<td>4/3</td>
<td>20 10 10</td>
<td>50 50 50</td>
</tr>
<tr>
<td>DS22</td>
<td>4/3</td>
<td>20 19 15</td>
<td>50 50 50</td>
</tr>
<tr>
<td>DS23</td>
<td>5/3</td>
<td>20 15 15</td>
<td>50 50 50</td>
</tr>
<tr>
<td>DS24</td>
<td>8/3</td>
<td>20 15 15</td>
<td>50 50 50</td>
</tr>
<tr>
<td>DS2-CBASP</td>
<td>3/2</td>
<td>20 10 10</td>
<td>50 50 50</td>
</tr>
</tbody>
</table>

Motivation

Methods and Algorithms

Results and Discussion

Experimental Results for Criterion 2

Discussion

Kun, Joelle, Susan

Active Learning for Developing Personalized Treatment
Experimental Results for Criterion 2

**DS24**

- **AARandom**
- **MINMAXPICS (SEQ)**
- **MINMAXPICS (GRP)**
- **Oracle Allocation (surrogate)**

**DS2-CBASP**

- **AARandom**
- **MINMAXPICS (SEQ)**
- **MINMAXPICS (GRP)**
- **Oracle Allocation (surrogate)**

---

Kun, Joelle, Susan

Active Learning for Developing Personalized Treatment
Related Work

- **RL**
  - action space is (subpopulation, treatment) pair
  - finite horizon (N)
  - goal is NOT maximizing cumulative reward

- **Budgeted Multi-armed Bandit Problem**: optimize a goal function constrained by a time or cost budget
  - pick an arm of a slot machine with maximal payoff
  - design a classifier with minimal prediction risk
  - estimate quantities with minimal variances
    (GAFS-MAX, Antos et al, 2008)
A minmax bandit model for characterizing the quality of a treatment rule

Potential in cost saving in comparison with a completely randomized exploration policy.

Optimization Criteria

- Why “max” or “uniformly good”? computational issue, patient/clinician’s perspective.
- What if there exist several equally good treatments?
- output one treatment per subpopulation, minimize maximal error of choosing $\delta$-bad treatment for prespecified $\delta$
- allow output multiple treatments per subpopulation, minimize maximal error of failing to exclude a “bad” treatment
Modeling choice. Bandit with covariate model, contextual bandits? How to quantify the quality of treatment rules for treating a particular patient?

Provide a way to estimate the required total budget $N$ in order to provide a high quality treatment rules.

use electronic medical record to discover biomarkers and recruit patients.