Active Learning for Personalizing Treatment

Kun Deng[1] Joelle Pineau[2] Susan Murphy[1]

[1]Department of Statistics
University of Michigan

[2]Department of Computer Science
McGill University

April 6, 2011

Outline

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
  - provide reliable evidence that informs clinical decision making
    - construct decision rules from clinical data that are tailored to individual heterogeneity
    - quantify confidence/uncertainty of these decision rules
  - make better use of clinical trial resources
    - number of recruitments

A Motivating Example

- Patients are categorized into subpopulations $c_1 \sim c_4$ based on biomarkers
- Two treatment actions a_1 and a_2
- An individualized treatment assignment 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 estimates of mean responses for subpopulation c_i

- The uncertainty in the estimated treatment effect lies in $\mathrm{Var}[\hat{\mu}_{i1} - \hat{\mu}_{i2}] = \mathrm{Var}[\hat{\mu}_{i1}] + \mathrm{Var}[\hat{\mu}_{i2}]$
- Further we want to treat all four subpopulations, so we need to control the uncertainty for all i.

Cont'd

- Current Practice
  - Recruit from the entire population as patients arrive: patients in the trial roughly reflect their natural composition.
  - A post subgroup analysis is used to calculate $\mathrm{Var}[\hat{\mu}_{i1} - \hat{\mu}_{i2}]$, often yielding highly variable responses for some subpopulations
  - Question: how to intelligently recruit patients from subpopulations in order to construct a uniformly-good treatment policy.

Cont'd

- Our Approach
  - a minimax bandit model that intelligently recruits patient from different subpopulations and assigns them to different treatments
  - minimize the largest variance of the estimated treatment effects among the different subpopulations
- 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 is extremely 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 μ_{ij} (corresponding to the primary outcome of action (i,j)) and variance σ_{ij}^2.
- Based on our goal of creating good ITRs, we want to control the maximum estimation loss for all subpopulations

A MiniMax Bandit Problem

Some Definitions

- The error of estimation for an arm: $L_{ij}^n = E[(\hat{\mu}_{ij}^n - \mu_{ij})^2] = \mathrm{Var}[\hat{\mu}_{ij}^n] = \frac{\sigma_{ij}^2}{T_{ij}}$,
- T_{ij} being the number of pulls for (i,j)
- The loss for a bandit : $L_i^n = \sum_{j=\{1,2\}} L_{ij}^n$
- The overall loss of an active learning policy π: $L^n(\pi) = \max_{1 \le i \le C} L_i^n$
- An oracle policy π^{oracle} that knows the variances σ_{ij}^2
- The excessive loss of π compared to an optimal oracle policy $L^n(\pi) - L^n(\pi^{\mathrm{oracle}})$, goal is to find π that minimizes it.

The Oracle Algorithm

The Oracle Algorithm

$$\begin{aligned} \underset{T_{ij}}{\text{minimize}} \quad & \max_i \sum_j \frac{\sigma_{ij}^2}{T_{ij}} \\ \text{s.t.} \quad & \sum_i \sum_j T_{ij} = N \\ & T_{ij} \geq 0 \end{aligned}$$

Solution

$$T_{ij}^* = \frac{\sigma_{ij} \sum_j \sigma_{ij}}{\sum_i (\sum_j \sigma_{ij})^2} N$$

$$r^* = \frac{\sum_i (\sum_j \sigma_{ij})^2}{N}$$

r^* is the optimal value of the objective function.

The Oracle Algorithm

The Oracle Algorithm

$$\begin{aligned} \underset{T_{ij}}{\text{minimize}} \quad & \max_i \sum_j \frac{\sigma_{ij}^2}{T_{ij}} \\ \text{s.t.} \quad & \sum_i \sum_j T_{ij} = N \\ & T_{ij} \geq 0 \end{aligned}$$

Solution

$$T_{ij}^* = \frac{\sigma_{ij} \sum_j \sigma_{ij}}{\sum_i (\sum_j \sigma_{ij})^2} N$$

$$r^* = \frac{\sum_i (\sum_j \sigma_{ij})^2}{N}$$

r^* is the optimal value of the objective function.

AREOA Algorithm

- $\left\{ \frac{\sigma_{ij} \sum_j \sigma_{ij}}{\sum_i (\sum_j \sigma_{ij})^2}; i \in \{1, \dots, C\}, j \in \{1, \dots, K\} \right\}$ forms a probability distribution
  - if an active learning algorithm samples according to this at each time point,
  - $E[T_{ij}(\pi)] = T^*_{ij}$
  - if σ_{ij} is large or $\sum_{.} \sigma_{i.}$ is large, arm (i,j) should be pulled more often
- Use plugin estimates $\hat{\sigma}_{ij}$ in the active learning policy
- Compared against uniformly random–AARandom and one related algorithm GAFS-MAX

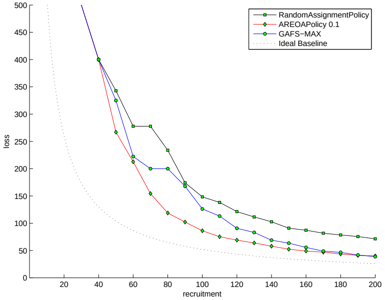
Experiments

- All arms were initialized with B=5 pulls
- AREOA uses ε-greedy strategy to keep a small probability of random exploration, $\varepsilon = 0.1$

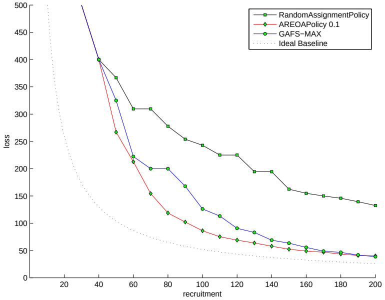
Table: datasets for the experiment

dataset	subpopulation/ treatments	distributions	means	variances
DS1	4/2	$\begin{pmatrix} .25 \\ .25 \\ .25 \\ .25 \end{pmatrix}$	$\begin{pmatrix} 1 & 4 \\ 2 & 2 \\ 4 & 1 \\ 2 & 2 \end{pmatrix}$	$\begin{pmatrix} 1000 & 1000 \\ 100 & 100 \\ 100 & 100 \\ 100 & 100 \end{pmatrix}$
DS2	4/2	$\begin{pmatrix} .1 \\ .3 \\ .3 \\ .3 \end{pmatrix}$	$\begin{pmatrix} 1 & 4 \\ 2 & 2 \\ 4 & 1 \\ 2 & 2 \end{pmatrix}$	$\begin{pmatrix} 1000 & 1000 \\ 100 & 100 \\ 100 & 100 \\ 100 & 100 \end{pmatrix}$
DS3	8/2	$\begin{pmatrix} .125 \\ .125 \\ \dots \\ .125 \end{pmatrix}$	$\begin{pmatrix} 2 & 2 \\ 2 & 2 \\ \dots & \dots \\ 2 & 2 \end{pmatrix}$	$\begin{pmatrix} 5 & 5 \\ 10 & 10 \\ \dots & \dots \\ 640 & 640 \end{pmatrix}$
DS4	4/2	$\begin{pmatrix} .25 \\ .25 \\ .25 \\ .25 \end{pmatrix}$	$\begin{pmatrix} 1 & 4 \\ 2 & 2 \\ 4 & 1 \\ 2 & 2 \end{pmatrix}$	$\begin{pmatrix} 100 & 1000 \\ 100 & 1000 \\ 100 & 1000 \\ 100 & 1000 \end{pmatrix}$
DS-CBASP	3/2	$\begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}$	$\begin{pmatrix} 10.9 & 16.2 \\ 9.3 & 19.4 \\ 12.9 & 15.8 \end{pmatrix}$	$\begin{pmatrix} 99.3 & 79.7 \\ 110.7 & 55.9 \\ 103.5 & 78.6 \end{pmatrix}$

Cont'd



DS1

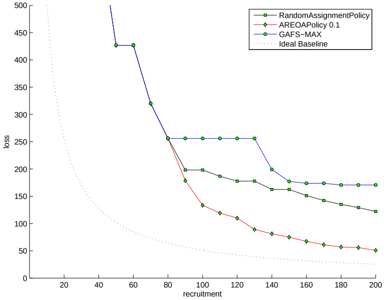


DS2

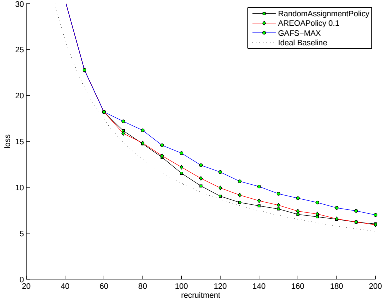
Table: datasets for the experiment

dataset	subpopulation/ treatments	distributions	means	variances
DS1	4/2	$\begin{pmatrix} .25 \\ .25 \\ .25 \\ .25 \end{pmatrix}$	$\begin{pmatrix} 1 & 4 \\ 2 & 2 \\ 4 & 1 \\ 2 & 2 \end{pmatrix}$	$\begin{pmatrix} 1000 & 1000 \\ 100 & 100 \\ 100 & 100 \\ 100 & 100 \end{pmatrix}$
DS2	4/2	$\begin{pmatrix} .1 \\ .3 \\ .3 \\ .3 \end{pmatrix}$	$\begin{pmatrix} 1 & 4 \\ 2 & 2 \\ 4 & 1 \\ 2 & 2 \end{pmatrix}$	$\begin{pmatrix} 1000 & 1000 \\ 100 & 100 \\ 100 & 100 \\ 100 & 100 \end{pmatrix}$
DS3	8/2	$\begin{pmatrix} .125 \\ .125 \\ \dots \\ .125 \end{pmatrix}$	$\begin{pmatrix} 2 & 2 \\ 2 & 2 \\ \dots & \dots \\ 2 & 2 \end{pmatrix}$	$\begin{pmatrix} 5 & 5 \\ 10 & 10 \\ \dots & \dots \\ 640 & 640 \end{pmatrix}$
DS4	4/2	$\begin{pmatrix} .25 \\ .25 \\ .25 \\ .25 \end{pmatrix}$	$\begin{pmatrix} 1 & 4 \\ 2 & 2 \\ 4 & 1 \\ 2 & 2 \end{pmatrix}$	$\begin{pmatrix} 100 & 1000 \\ 100 & 1000 \\ 100 & 1000 \\ 100 & 1000 \end{pmatrix}$
DS-CBASP	3/2	$\begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}$	$\begin{pmatrix} 10.9 & 16.2 \\ 9.3 & 19.4 \\ 12.9 & 15.8 \end{pmatrix}$	$\begin{pmatrix} 99.3 & 79.7 \\ 110.7 & 55.9 \\ 103.5 & 78.6 \end{pmatrix}$

Cont'd



DS3



DS-CBASBP

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)

Summary

- A minmax bandit model for characterizing the quality of a treatment rule
- Potential in cost saving in comparsion with a completely randomized exploration policy.
- Modeling choice. Provide a way to estimate the required total budget N in order to provide a high quality treatment rules.
- Other criteria for defining "good" treatment rules.
    - maximal error of confusing suboptimal treatment with the true best treatment for any subpopulation
    - maximal total number of correctly identified "best" treatment
- use electronic medical record to discover biomarkers and recruit patients.