Bayesian Additive Regression Trees (BART) and Precision Medicine

Brent Logan
Medical College of Wisconsin
Joint work with Prakash Laud, Rob McCulloch, and Rodney Sparapani

Sep 30, 2017
Background

- Patients are heterogeneous and may respond differently to treatment

- **Goal of Precision Medicine** Identify which patients respond best to which treatment and tailor treatment to individual patients

- Personalization based on patient clinical features, biomarkers, genetic information

- Individualized treatment rule (ITR): providing a therapy with the best predicted outcome for that individual

- Extension of subgroup analysis

- Such personalized therapy can improve population health measures
Strategies for obtaining an ITR

- Policy search: directly optimize an estimator of the expected outcome of a treatment rule by searching over a class of rules.
- Predictive modeling of patient outcome
  - Good prediction accuracy needed to ensure good performance of ITR
  - Flexible prediction models to handle potentially complex interactions between treatment and covariates
Background

Why BART?

- Excellent performance as a flexible prediction model
- Natural quantification of uncertainty to assess the benefit of individualized treatment
Background

- **Notation:**
  - $Y$: binary outcome of interest (higher values are desired),
  - $A = \{-1, 1\}$: treatment
  - $X$: covariates of interest (biomarkers, clinical characteristics)

- **Individualized Treatment Rule (ITR), $g(x)$:**
  Treatment rule $g(X)$ is a map from the domain of $X$ to $A$, so that a patient with covariate $X$ is recommended treatment $g(X)$. 
Background

- **Value function**: expected outcome if all patients were treated according to the rule \( g \),

\[
V(g) = E[ E(Y|X, A = g(X)) ].
\]

- Measures population impact

- **Optimal ITR** \( g_0 \): satisfies \( V(g_0) \geq V(g) \forall g \).

- This is true if \( g_0(x) = \arg \max_a E(Y|X, A) \)

- Assign each patient the treatment which has the highest expected outcome.
BART Individualized Treatment Rule (ITR)

- **BART model**

\[ p(Y = 1|x, a, f) = \Phi(\mu_0 + f(x, a)), \]

where \( f \) is expressed as the sum of trees.

- \( f \) is viewed as the underlying parameter.

- Each MCMC sample results in draws \( f_d, d = 1, \ldots, D \) from the function \( f \).

- Optimal ITR: choose value of \( a \) which maximizes

\[ E(Y|x, a) = p(Y = 1|x, a). \]
BART Individualized Treatment Rule (ITR)

- Posterior predictive distribution integrated over $f$,

$$p(Y = 1|x, a) = \int p(Y = 1|x, a, f) dP(f|\text{Data}).$$

- Use MCMC samples to approximate this integral

$$p(Y = 1|x, a) \approx \frac{1}{D} \sum_{d=1}^{D} p(Y = 1|x, a, f_d) \equiv \bar{p}(x, a).$$

- BART ITR:

$$g_{\text{BART}}(x) = \arg \max_a \bar{p}(x, a).$$
Inference on value of any ITR $g$

- Value function of an ITR $g$ is a function of $f$ given by

$$V(g, f) = E_X[p(Y = 1|x, g(X), f)].$$

- Posterior samples of value function $V_d(g)$ given by plugging in draws of $f$,

$$V_d(g) = V(g, f_d).$$

- Expectation w.r.t $X$ often done by averaging over observed covariate distribution.

- These quantify uncertainty about the value function of $g$
Inference on value of Optimal ITR

- Optimal ITR is a function of $f$
- Given $f$, optimal action is
  \[ a(x, f) = \arg \max_p p(Y = 1|x, a, f), \]
  with corresponding maximum success probability
  \[ p^*(x, f) = \max_a p(Y = 1|x, a, f). \]
- Value function of optimal ITR $g_0$ is also a function of $f$
  \[ V^*(g, f) = E_x[p^*(x, f)]. \]
- Posterior samples of value function $V^*(g)$ given by plugging in draws of $f$,
  \[ V^*_d(g) = V^*(g, f_d). \]
- These quantify uncertainty about the value of the (random) Optimal ITR.
Simulation settings

- ITR’s generated using training dataset with $n = 500$

- **Setting I**: Similar to [XYZ+15] but reduced treatment interaction term. 5 additional binary covariates $X_A : X_E$, 5 ordinal covariates $X_a : X_e$ with four categories, and one or two continuous covariates $X_{Ca}, X_{Cb}$.

- **Setting II**: Identical to [KJH14]. Up to 3 independent continuous markers $X_1, X_2, X_3$.

- Wide range of settings with no interaction, linear interaction, nonlinear interaction, varying link functions.
Simulation metrics

- Each ITR applied to a fixed independent test dataset of size 2000 to determine the value function.
- Average value function across 50 replicate training sets used to compute the fraction of each ITR value function relative to the true optimal value function.
Simulations

- **Cross-validation**: Used to select number of trees \((m = 80, 200)\) and \(k\) parameter \((0.2, 0.8, 2.0)\).

- **Competing Methods**:
  - Regularized Outcome Weighted Subgroup Identification (ROWSI) \([XYZ^{+}15]\)
  - Outcome Weighed Learning (OWL): \([ZZRK12]\)
  - Random forest (RF) for outcome prediction with cross validation of number of trees and minimum node size
  - Boosting with classification tree working model (KANG): \([KJH14]\)
Simulation results I

![Graph showing simulation results for different methods across various scenarios.](image)

- **Method**
  - BARTcv
  - BARTd
  - KANG
  - OWL
  - RF
  - ROWSI

- **Scenarios**
  - A2, B2, C2, D2, E2, F2, G2, H2

- **Y-axis**
  - Fraction of optimal Value

- **X-axis**
  - Scenario

The graph illustrates the performance of different methods across various scenarios, with the fraction of optimal value plotted against scenario.
Operating Characteristics

- Features demonstrated with $n = 500$ (left) and $n = 5000$ (right) training set sample sizes using either (1) complex treatment interaction model,

$$P(Y = 1|A, X) = [1 + \exp \{-0.1 - 0.2X_1 + 0.2X_2 - 0.1X_3 + 0.5X_1^2 \\
+ A(-0.5 - 0.5X_1 - X_2 - 0.3I(X_3 > 0.5) + 0.5X_1^2)\}]^{-1}.$$ 

and (2) no treatment interaction model

$$P(Y = 1|A, X) = [1 + \exp\{-0.1 - 0.2X_1 + 0.2X_2 - 0.1X_3 + 0.5X_1^2 - 0.3A\}]^{-1}.$$ 

- Single dataset predictions of posterior mean treatment differences vs. truth (top)

- Repeated data simulation results: bias (bottom), coverage of posterior intervals for value function was 90% for $n = 500$ and 95% for $n = 5000$. 
Operating characteristics: Complex interaction
Operating characteristics: No interaction
Summarizing the BART ITR

- ITR based on BART does not directly yield a simple interpretable rule.
- Separate modeling of outcome from determination of interpretable rule, by developing an approximation to BART ITR which is interpretable and has good performance.
- “Fit-the-fit” strategy, develop a single tree fit to the posterior mean treatment differences (Data) as a function of patient characteristics.
- Quality of such an approximation can be assessed using $R^2$ between single tree and BART prediction model.
BMT Example

- **Cohort:** 3802 patients receiving reduced intensity hematopoietic cell transplant between 2011-2013 for a variety of hematologic malignancies, with data reported to the Center for International Blood and Marrow Transplant Research.

- **Patient, donor, and disease factors:** age, race/ethnicity, performance score, CMV status, disease, remission status, disease subtypes, chemosensitivity, interval from dx to tx, donor type, HLA matching between donor and recipient, prior autologous tx, gender matching between donor/recipient, comorbidity score, year of tx

- **Treatment of interest:** Conditioning regimen used (Flu/Mel vs. Flu/Bu)

- Observational cohort well balanced between regimens indicating some equipoise
BMT Example: individual patient inference

- **Outcome:** 1 year survival (binary due to minimal censoring <1yr).
- Fitting of BART model provides samples from $p(Y = 1|x, a, f_d)$ for $d = 1, \ldots, D$.
- Patient specific difference in 1 yr survival:
  $p(Y = 1|x, a = \text{Flu/Mel}, f_d) - (Y = 1|x, a = \text{Flu/Bu}, f_d)$
- Plot the posterior mean of these differences for each patient, sorted by the magnitude of the difference. Inter-quartile ranges and 95% posterior intervals shown.
- Plot posterior probability that Flu/Mel has higher 1yr survival than Flu/Bu:
  $$\frac{1}{D} \sum_{d} I(P(Y = 1|x, a = \text{Flu/Mel}, f_d) > P(Y = 1|x, a = \text{Flu/Bu}, f_d))$$
BMT Example: individual patient inference

Waterfall plots

Posterior mean difference in 1 yr survival

Patient Number (Ordered)

Posterior probability of FluMel superiority

Patient Number (Ordered)
BMT Example: population inference and value functions
BMT Example: Explaining to physicians

Difference in 1 yr. survival (Flu/Mel−Flu/Bu)

- NHL subtype = Aggressive B-cell
  - Other
  - AML
  - AML, Other

- NHL subtype = Other
  - AML

0.016
n=3802
[ −0.075 , 0.062 ]

−0.054
n=676
[ −0.088 , −0.027 ]

−0.07
n=235
[ −0.089 , −0.048 ]

−0.046
n=441
[ −0.074 , −0.026 ]

0.0092
n=1573
[ −0.002 , 0.017 ]

0.053
n=1553
[ 0.036 , 0.064 ]
Summary

- Excellent performance of BART for defining ITR’s which optimize patient outcomes.
- Provides direct inference on the value function of the ITR through posterior samples.
- Post processing of BART inference can provide approximations which have good performance and are clinically interpretable.
