Targeting for Long-Term Outcomes

Jeremy Yang[1], Dean Eckles[1], Paramveer Dhillon[2], Sinan Aral[1]

[1]Massachusetts Institute of Technology

[2]University of Michigan

November 2020

Learning to target interventions

- We have some intervention — a discount, an ad campaign, or a piece of content to recommend
- How should we decide who to target with it?
- We will use randomized experiments to generate data about for whom it might work.
- A policy and its value:

$$\pi : X \to \Delta A$$

$$\mathbb{E}[Y_i] := \mu(x_i, \pi(x_i))$$

$$V(\pi) := \frac{1}{n} \sum_i \mathbb{E}[Y_i]$$

Long-term outcomes

- What makes long-term outcomes different?
- When primary outcome of interest is short-term, a policy can be directly optimized on it before treating the next batch.
- But long-term outcomes are not observed in the short-term.
  - Just wait. But can't learn or take actions in between.
  - Optimize a short-term proxy instead. But it might be myopic and not well aligned with the long-term outcome.

Empirical context: The Boston Globe

- Boston Globe is the largest newspaper in New England.
- $6.93/week billed every 4 weeks for a total of $27.72.
- How can we retain subscribers and maximize their long-term revenue by targeting discounts?
- Two rounds of experiments on all digital subscribers.
- What we observe for each customer:
  - demographics (e.g., zip code)
  - account activities (e.g., billing address change, payment decline, credit card expiration date, complaints)
  - content consumption (what articles and when)
  - cancellation and revenue

Our approach

- Use statistical surrogacy to impute the missing long-term outcomes and optimize the policy on imputed outcomes.
- Implement the optimal policy via bootstrapped Thompson sampling to account for non-stationarity.[1]
- Application: Who should I give discount to, and how much, to maximize total subscription revenue over the next 3 years?

[1] Add randomness in every round of targeting to never stop exploring.

Framework

- Imputation of long-term outcome: $\tilde{Y}$ in place of Y.
- Policy learning:

$$\pi^* = \text{argmax}_{\pi \in \Pi} \tilde{V}(\pi)$$

- Policy implementation: adapt to potential non-stationarity by continuing to experiment π_t^*

Imputation of long-term outcome

- Statistical surrogate: a short-term proxy variable S that if conditioned on, makes long-term outcome Y independent of treatment A (Prentice 1989).
- One causal model where this is satisfied: S fully mediates the treatment effect from A to Y.
- Perhaps more plausible if S is detailed (potentially high-dimensional). Can find a set of surrogates S that jointly mediate the treatment effect and learn a surrogate index (Athey et al., 2020).

Imputation of long-term outcome

- Suppose we have two datasets: experimental (e) and historical/observational (h).
- In experimental data we observe (x_i, a_i, s_i).
- In historical data we observe (x_j, s_j, y_j).
- The long-term outcome in the experimental data can be imputed as:
$$\tilde{Y}_i = \mathbb{E}_h[Y|s_i, x_i]$$
- $\tilde{Y}_i$ is called surrogate index (SI) (Athey et al., 2020).

Imputation of long-term outcome

- As in Athey et al. 2020
- Assumption (1): unconfoundedness and positivity.

$$A_i \perp\!\!\!\perp (Y_i(a), S_i(a)) \; \forall a \in \mathbb{A}, i \in e$$
$$0 < e_a(x) < 1 \; \forall a \in \mathbb{A}, x \in \mathbb{X}$$

- Assumption (2): surrogacy or mediation.

$$A_i \perp\!\!\!\perp Y_i \mid S_i, X_i, i \in e$$

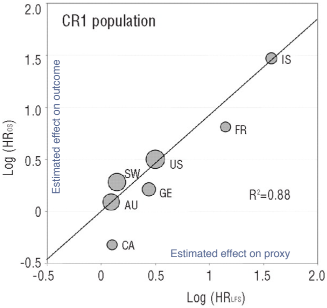
- Assumption (3): comparability.

$$Y_i \mid S_i, X_i, i \in e \sim Y_j \mid S_j, X_j, j \in h$$

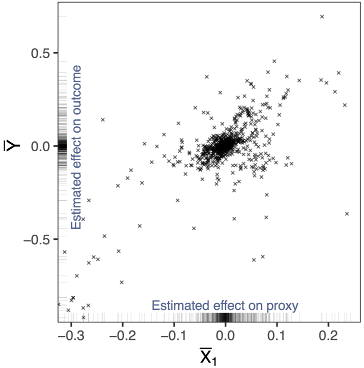
Imputation of long-term outcomes

- Under assumptions (1)-(3), the treatment effect on surrogate index $\tilde{Y}$ recovers the treatment effect on Y.
- Actually can work with a weaker sign-preserving assumption
- We prove that $\tilde{Y}$ can also be used for policy learning: optimal policy learned on $\tilde{Y}$ recovers the optimal policy learned on Y.

Surrogates: Evaluation with experiments



Buyse, M., et al. (2011). Leukemia-free survival as a surrogate end point for overall survival in the evaluation of maintenance therapy for patients with acute myeloid leukemia in complete remission. *Haematologica*, 96(8), 1106-1112.



Policy learning with doubly-robust scores

- Doubly robust scores for each user-action pair (Dudik et al., 2014; Athey and Wager, 2017; Zhou et al., 2018):

$$\hat{\gamma}_a(x_i) = \hat{\mu}(x_i, a) + \frac{Y_i - \hat{\mu}(x_i, a)}{e_a(x_i)} \cdot 1_{\{a_i = a\}}$$

- Binary cost-sensitive classification (multi-action):

$$\pi^*(x_i) = \text{argmax}_\pi \ \frac{1}{n} \sum_i \left(\hat{\gamma}_1(x_i) - \hat{\gamma}_0(x_i)\right) \cdot (2\pi(x_i) - 1)$$

- More efficient than outcome regression or causal forest.

Policy implementation

- Implement π^* via bootstrapped Thompson sampling (Eckles and Kaptein, 2014;Osband et al., 2015, 2017).
- Estimate a π^*_b via cost-sensitive classification for each bootstrap sample $b \in B$, then average them.
- $\pi^*(A = a|x_i) = \frac{1}{|B|} \sum_b \mathbb{1}_{\{\pi^*_b(A=a|x_i)\}}$
- Account for non-stationarity and can iterate on future cohorts.

Two field experiments

- First cohort:
  - 2018/8, 45K users, 1K treated.
  - treatment: $4.99/week for 8 weeks.
- Second cohort:
  - 2019/7, 95K users, 6K treated.
  - treatment: thank you email only, $20 gift card, $5.99/week for 8 weeks, $5.99/4, $4.99/8, $3.99/8.
- No overlaps of treated users between two cohorts.

Treatments

The Boston Globe

Thank you.

Dear Jessie,

Thank you for being a loyal Globe.com subscriber. As a Boston Globe reader, your support allows us to continue producing award-winning journalism.

To say thank you, we are reducing your subscription rate to $4.99/week for the next 8 weeks. This discounted rate will be automatically applied to your next two invoices.

Go to Globe.com →

The Boston Globe

Thank you.

Dear David,

As a thank you for supporting journalism and placing your trust in us, we are reducing your subscription rate to $5.99/week for the next four weeks. To claim this discounted rate, please click the button below and the discount will be automatically applied to your subscription.

As a Globe subscriber, you are a critical part of a community that cares deeply about the issues that affect not only our region but the world. The Globe is committed to continuing to spark inspiring and informing in-depth discussions on the most important issues that will shape the future of our community. I urge you to add your voice by commenting on articles or by joining us at community events. Your ongoing subscription directly fuels this mission and I write today to let you know that we don't take for granted your decision to support Pulitzer Prize-winning journalism in New England.

Sincerely,

John W. Henry
Publisher

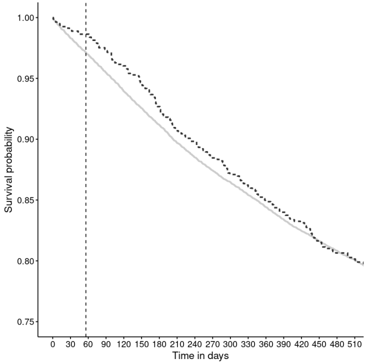
Claim My Discount →

Claim discount by 7/16. Offer expires at midnight on 7/16.

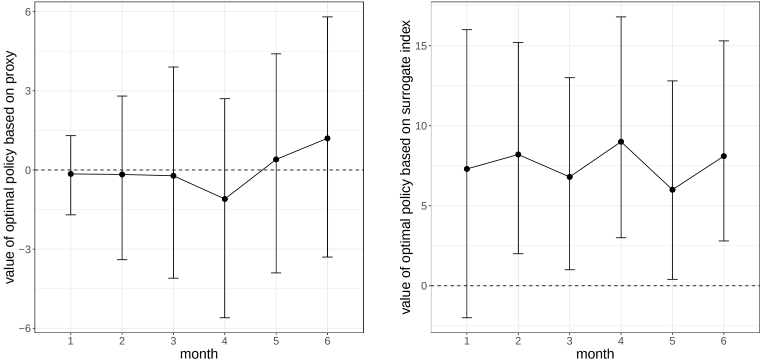
Procedure

- Step (1): implement a behavior/design policy π_0 on the first cohort that balances exploration and exploitation. detail
- Step (2): observe surrogates S (1-6 month revenue and content consumption) and impute 3-year revenues $\tilde{Y}(s_i, x_i)$ on the first cohort using historical data. detail
- Step (3): learn π^* with $\tilde{Y}$ and implement it via bootstrapped Thompson sampling on the second cohort. detail
- Step (4): repeat step (2) and (3) to update the optimal policy after treating each cohort to account for non-stationarity.

Result: first cohort (survival curve)

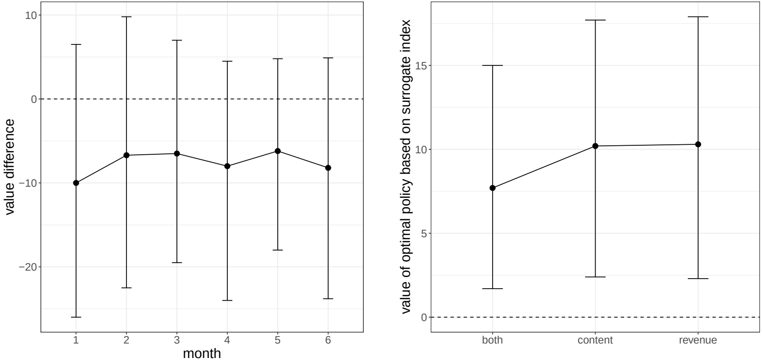


Validation of surrogate index



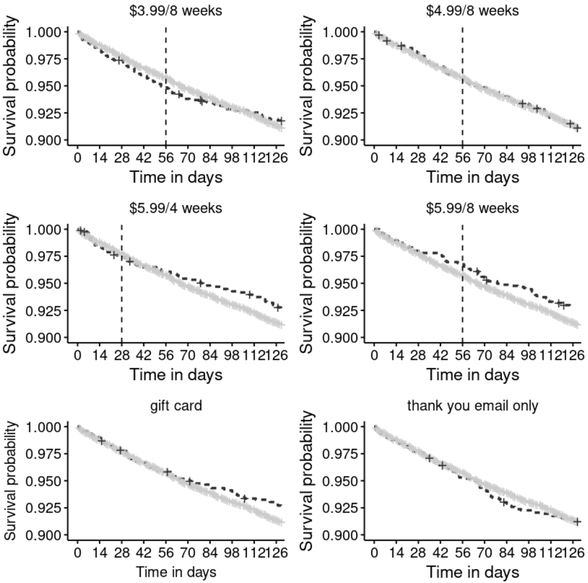
- Left: The value of optimal policy learned on short-term proxies (1-6 month revenue), they don't outperform the status quo.
- Right: The value of optimal policy learned on surrogate indices constructed with surrogates from 1-6 months, they outperform the status quo (except for month 1).

Validation of surrogate index



- Left: The value difference between optimal policies learned on surrogate indices and true outcome, they are statistically indistinguishable.
- Right: Comparing surrogate indices constructed using different variables: content consumption only, short-term revenue only, and both, there is no significant differences.

Result: second cohort (survival curve)



Result: second cohort (optimal policy)

condition	percentage
control	23%
thank you email only	25%
gift card	$< 1\%$
\$5.99/8 weeks	25%
\$5.99/4 weeks	27%
\$4.99/8 weeks	$< 1\%$
\$3.99/8 weeks	$< 1\%$

- Checking for non-stationarity: covariate shift and concept shift

Conclusion

- We combine a surrogate index with policy learning to optimize targeting for long-term outcomes.
- Prove that the approach works under certain assumptions.
- Implement in adaptive randomized experiments.
- Validate it empirically by targeting discount to digital subscribers of Boston Globe.
- Increases 1.5-year and 3-year revenue by $15 and $40 per user relative to the status quo in the two cohorts.
- 3-year total revenue impact sums up to $4-5 million.

Thanks

- Coauthors: Dean Eckles, Paramveer Dhillon, Sinan Aral.

Some personal advertising:

- I'm on the academic job market.
- General interest: optimizing managerial decisions with causal inference, machine learning and unstructured data (e.g., video, text and network).

- Email: jeremy.z.yang@gmail.com
- Website: jeremyzyang.github.io
- Twitter @jeremyzyang

Related work

- Applying doubly robust scores based policy evaluation and learning, more efficient than e.g., outcome model, causal forest (vs. Hitsch and Misra, 2018; Simester et al., 2019 (a,b))
- Using bounded outcome to design a starting policy that's better than uniformly at random with minimal assumption
- Dynamic policy that responses to changes in the environment (vs. Athey and Wager, 2017; Zhou et al., 2018; Hitsch and Misra, 2018; Simester et al., 2019 (a,b))
- Learning a policy to optimize long-term outcomes with surrogate index (vs. Athey et al., 2019)

Other related work

- Methodological:
  - Contextual bandit (Langford and Zhang, 2008; Hauser et al., 2009; Li et al., 2011; Burtini et al, 2015; Schwartz et al., 2017; Misra et al., 2019)
  - Off-policy evaluation and optimization (Dudik et al., 2014; Li et al., 2011; Athey and Wager, 2017; Hitsch and Misra, 2018; Simester et al., 2019 (a,b); Zhou et al., 2018)
- Substantive:
  - churn management (Ascarza et al., 2016; Ascarza et al., 2017; Ascarza 2018; Godinho de Matos et al., 2018)

back

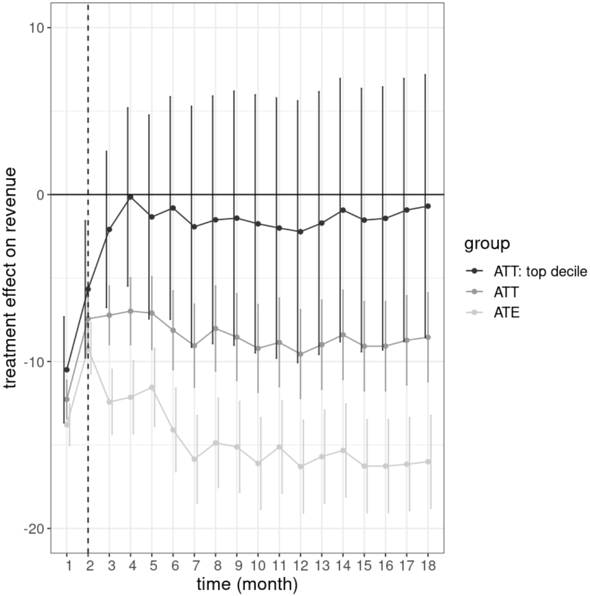
Supplementary Materials

Design policy for first cohort

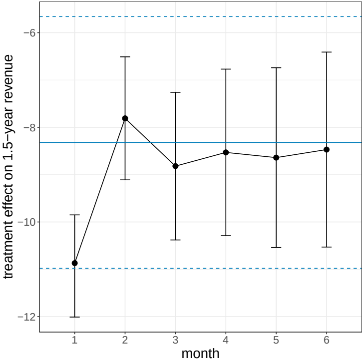
- The choice of π_0 can encode prior knowledge.
- Bounded outcome imposes restrictions on the magnitude of treatment effect with minimal assumptions (e.g., monotonicity: treatment doesn't increase churn risk).
- Y is long-term (e.g., 3 year) churn probability, which is bounded between 0 and 1.
- We first train a classifier churn prediction to predict $Y(0)$ and then treat users with higher $Y(0)$ with higher probability. simulation
- $0 < \pi_0(a|X_i) < 1$ (some practical level of positivity).

back

Result: first cohort (treatment effect)



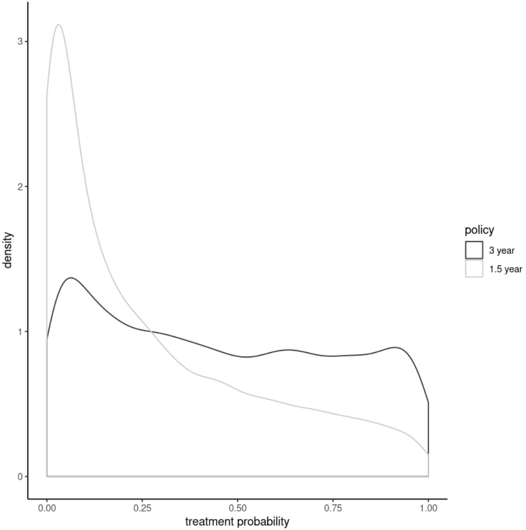
Validation of surrogate index



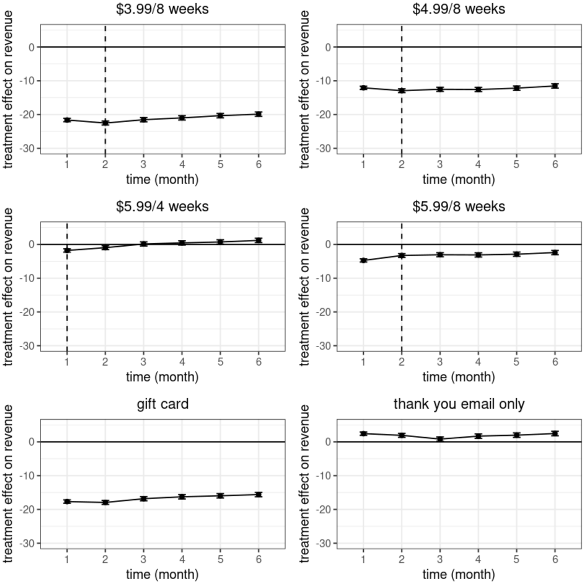
Average treatment effect on the treated (ATT) on revenue using surrogates from the first 6 months, blue line is the ATT estimated using true 1.5-year revenue.

Learned policy plus exploration

Add exploration with bootstrapped Thompson sampling



Result: second cohort (treatment effect)



Proof

- $\tilde{Y}$ is valid for policy evaluation:

$$
\begin{aligned}
V(\pi) &= n^{-1}\sum_{i}^{n}\{\pi(x_i)Y_i(1) + (1-\pi(x_i))Y_i(0)\} \\
&= n^{-1}\sum_{i}^{n}\{\pi(x_i)\mathbb{E}[\frac{A_iY_i}{e(x_i)}] + (1-\pi(x_i))\mathbb{E}[\frac{(1-A_i)Y_i}{1-e(x_i)}]\} \\
&= n^{-1}\sum_{i}^{n}\mathbb{E}[\pi(x_i)\frac{A_iY_i}{e(x_i)} + (1-\pi(x_i))\frac{(1-A_i)Y_i}{1-e(x_i)}] \\
&= n^{-1}\sum_{i}^{n}\mathbb{E}[\mathbb{E}[\pi(x_i)\frac{A_iY_i}{e(x_i)} + (1-\pi(x_i))\frac{(1-A_i)Y_i}{1-e(x_i)}|s_i,x_i]] \\
&= n^{-1}\sum_{i}^{n}\mathbb{E}[\pi(x_i)\frac{\mathbb{E}[A_i|s_i,x_i]\mathbb{E}[Y_i|s_i,x_i]}{e(x_i)} + (1-\pi(x_i))\frac{\mathbb{E}[1-A_i|s_i,x_i]\mathbb{E}[Y_i|s_i,x_i]}{1-e(x_i)}] \\
&= n^{-1}\sum_{i}^{n}\mathbb{E}[\pi(x_i)\frac{A_i\tilde{Y}_i}{e(x_i)} + (1-\pi(x_i))\frac{(1-A_i)\tilde{Y}_i}{1-e(x_i)}]
\end{aligned}
$$

back

Proof

- $\tilde{Y}$ is valid for policy learning.
- First, $\tilde{Y}$ recovers CATE on Y (Athey et al., 2019).

$$
\begin{aligned}
\tau(x_i) &= \mathbb{E}[Y_i(1) - Y_i(0)|x_i] \\
&= \mathbb{E}[\frac{A_i Y_i}{e(x_i)} - \frac{(1-A_i)Y_i}{1-e(x_i)}|x_i] \\
&= \mathbb{E}[\mathbb{E}[\frac{A_i Y_i}{e(x_i)} - \frac{(1-A_i)Y_i}{1-e(x_i)}|s_i, x_i]] \\
&= \mathbb{E}[\frac{\mathbb{E}[A_i|s_i, x_i]\mathbb{E}[Y_i|s_i, x_i]}{e(x_i)} - \frac{\mathbb{E}[1-A_i|s_i, x_i]\mathbb{E}[Y_i|s_i, x_i]}{1-e(x_i)}|x_i] \\
&= \mathbb{E}[\frac{A_i \tilde{Y}_i}{e(x_i)} - \frac{(1-A_i)\tilde{Y}_i}{1-e(x_i)}|x_i]
\end{aligned}
$$

- It follows immediately that policy learned on $\tilde{Y}$ recovers the optimal policy learned on Y. The two policies always make the same decisions (and therefore have the same value) as long as the CATE on $\tilde{Y}$ have the same sign as the CATE on Y.

Imputing long-term outcome with SI

- Y = 3-year revenue
- S = {short term (1-6 months) revenue, content consumption}
- Use a historical data set from 2015-2018 to estimate 3-year revenue as a function of S and X: $\mathbb{E}_h[Y|s_i, x_i]$
- Plug in the S and X from first cohort to impute the 3-year revenue for users in the experiment: $\tilde{Y}(s_i, x_i)$
- Use $\tilde{Y}(s_i, x_i)$ for policy learning

back

Efficient policy learning and implementation

- Estimate doubly robust scores for each user action pair in the first cohort via cross-fitting.
- Use doubly robust scores to construct label and case weight and use cost-sensitive classification to learn the optimal policy on the training set (80/20 split).
- Implement the optimal policy by repeating the procedure above on each bootstrap sample of the training set, and assigning each user to treatment with probability equal to the fraction of times that user is treated across all bootstrap samples (we clip probabilities near 0 and 1 to ensure that positivity assumption holds).
- When estimating π^* for future cohorts can pool data from the first two cohorts but weigh them by recency (Russa et al., 2019). Intuition: apply equal weighting across rounds when the environment is stable, only use data from the most recent round if environment is changing fast.

back

CATE

- CATE is the core of policy learning:

$$\begin{aligned}\max_\pi V(\pi) &= \max_\pi \; n^{-1}\sum_i Y(x_i, \pi(x_i)) \\ &= \max_\pi \; n^{-1}\sum_i \{\pi(x_i)Y_i(1) + (1-\pi(x_i))Y_i(0)\} \\ &= \max_\pi \; n^{-1}\sum_i \{\pi(x_i)(Y_i(1) - Y_i(0)) + Y_i(0)\} \\ &= \max_\pi \; n^{-1}\sum_i \{\pi(x_i)\tau(x_i) + Y_i(0)\}\end{aligned}$$

- $\pi^*(x_i) = 1 \iff \tau(x_i) > 0$

back

Policy evaluation

- Example: π_0 is a uniform policy, π_P only treats users who subscribed in the last 3 months.
- A doubly robust approach to estimating the value of counterfactual policy π_P (Dudik et al., 2014):

$$\hat{V}_{\mathrm{DR}}(\pi_P) = \frac{1}{n}\sum_i \left(\hat{\mu}(x_i, \pi_P) + \frac{\pi_P(A = a_i|x_i)}{\pi_0(A = a_i|x_i)} \cdot (Y_i - \hat{\mu}(x_i, a_i))\right)$$

$$\hat{\mu}(x_i, \pi_P) := \sum_{a \in A} \pi_P(A = a|x_i)\hat{\mu}(x_i, a)$$

Multi-action policy learning

- The doubly robust scores are the same:

$$\hat{\gamma}_a(x_i) = \hat{\mu}(x_i, a) + \frac{Y_i - \hat{\mu}(x_i, a)}{e_a(x_i)} \cdot 1_{\{a_i = a\}}$$

- Now the objective function is:

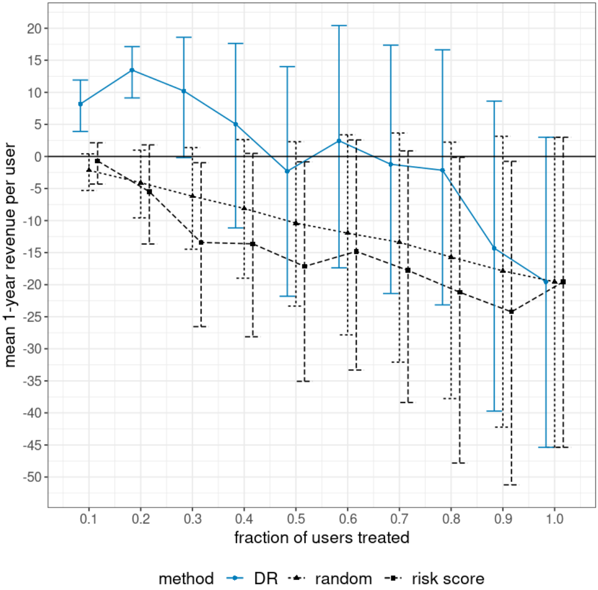
$$\pi^*(x_i) = \text{argmax}_{\pi \in \Pi} \; n^{-1} \sum_i < \hat{\gamma}(x_i), \pi(x_i) >$$

$< \cdot >$ is the dot product of doubly robust and policy vectors

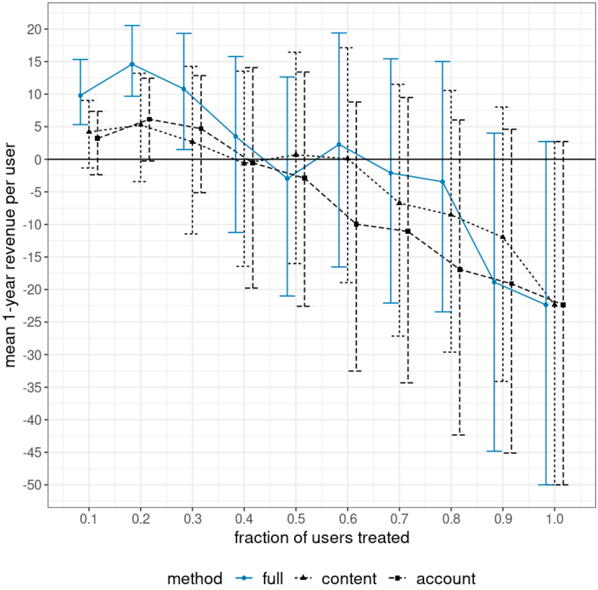
- It can also be reduced to a binary cost-sensitive classification problem by fitting a classifier for each pair of actions and then the optimal action is chosen by majority vote

back

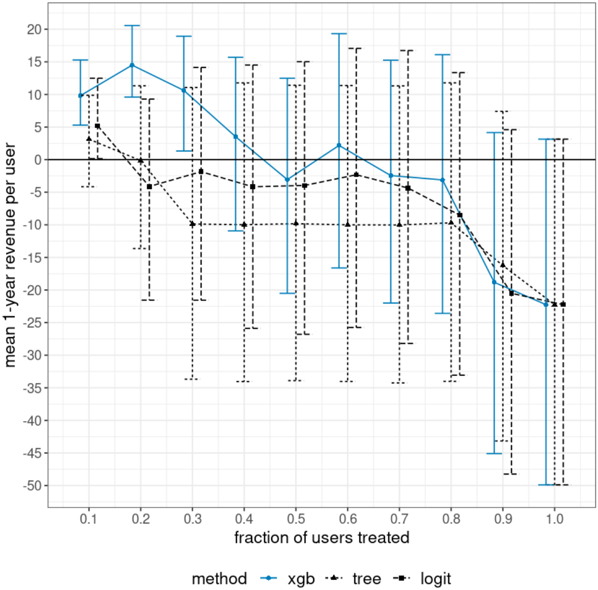
Result: first cohort (benchmark)



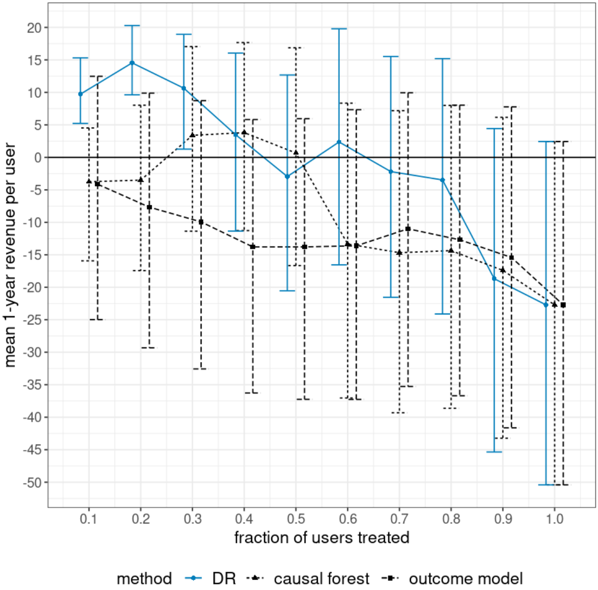
Result: first cohort (value of information)



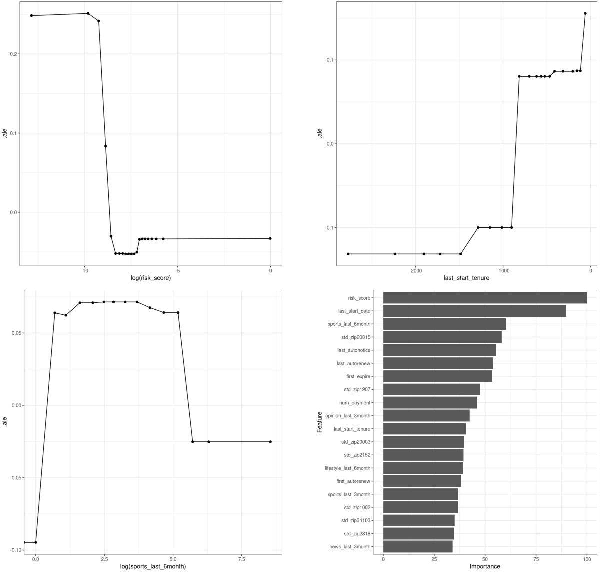
Result: first cohort (value of model)



Result: first cohort (value of model)



Result: first cohort (policy interpretation)



Potential and observed outcomes

- Potential outcomes:

$$Y_i(a) \text{ for } a \in \mathbb{A} \text{ (e.g., 0, 1)}$$

- We observe $Y_i = Y_i(A_i)$. In the special case of a binary action (i.e. treatment), that is

$$Y_i = A_i Y_i(1) + (1 - A_i) Y_i(0)$$

ATEs with fancier randomization

- We might have randomized people to treatment with different probabilities that depend on their characteristics.
- If these probabilities π_0 are known and not 0, we can use estimators that reweight using their inverse
- Normalized (Hàjek) estimator

$$\hat{\mu}_1^{\text{Hàjek}} = \frac{1}{\sum_i A_i} \sum_{i=1}^{N} Y_i \frac{A_i}{\pi_0(A_i|X_i)}$$

- Can estimate τ with, e.g., $\hat{\tau}^{\text{Hàjek}} = \hat{\mu}_1^{\text{Hàjek}} - \hat{\mu}_0^{\text{Hàjek}}$.

Policies

- A policy (or treatment rule, or treatment regime) determines how units are assigned to treatment.
- A homogeneous deterministic policy would assign all units to the same treatment, so we could write z_π to designate this treatment.
- Deterministic policies where not all units are given the same treatment assignment. Then we can designate the treatment that policy π assigns for unit i with $z_{\pi i}$ or the $\pi(a|X_i)$.
- Other cases:
  - Dynamic treatment regimes, reinforcement learning.
  - Stochastic policies that rely only on random samples from data about units.

Policy evaluation

- What is the mean outcome under some policy π?

$$\mu_\pi = [Y_i(z_{\pi i})]$$

- If π is stationary and possibly stochastic, it could be useful to write:

$$\mu_\pi = \sum_{z \in \mathbb{Z}} [Y_i(a)] \pi(a|X_i)$$

  This also makes clear the homogeneity of the policy for any units i and j such that $X_i = X_j$.

Policy evaluation from a Bernoulli experiment

How can we estimate μ_π from experimental data?

- For now assume Bernoulli experiment with $\Pr(A_i = 1) = \Pr(A_i = 0) = 1/2$
- For boring policy that would **assign everyone to treatment**:

$$\mu_1 = [Y_i(1)]$$

so can estimate this with

$$\hat{\mu}_1 = \frac{1}{\sum_i A_i} \sum_i Y_i A_i$$

This is just the sample mean in treatment.

Who should be targeted with an intervention?

- For each unit (e.g., customer), we observe their characteristics (i.e., features, covariates, context) $X_i \in \mathbb{X}$ and can choose an action $a \in \mathbb{A}$
- We then observe the outcome (i.e. reward) for that action $Y_i(a)$
- A *policy* π is a way of making these choices about actions. It maps from characteristics to actions, i.e., $\pi : \mathbb{X} \rightarrow \Delta(\mathbb{A})$
- Then targeting is a matter of finding a good (or the best) policy within some set of (perhaps simple) possible policies π, i.e.,

  $$\pi^* = \text{argmax}_\pi V(\pi)$$

  where $V(\pi) =_\pi [Y_i(A_i)]$
- We do this using past data, ideally where we've randomized the actions taken in some way

Policy evaluation from a Bernoulli experiment

- For concreteness, consider a policy that assigns treatment above some threshold on a scalar covariate. That is, units are treated iff $X_i > c$.

$$\pi_c(z|x) = (z = (x > c))$$

- Then to estimate μ_{π_c} we look at outcomes for units with $X_i > c$ that were treated and units with $X_i \leq c$ that weren't treated.

$$\hat{\mu}_{\pi_c} = \frac{1}{\sum_i (A_i = (X_i > c))} \sum_{i=1}^{N} Y_i \, (A_i = (X_i > c))$$

Policy evaluation with a design policy

- What if units were not assigned to treatment with equal probabilities, but rather according to a policy π_0?
- We assume conditional unconfoundedness and positivity, i.e., $Y_i(z) A_i | X_i$ and $0 < \pi_0(z|X_i) < 1$ for all $z \in \mathbb{Z}$ and x in the support of X_i.
- Need to account for heterogeneous probabilities of assignment to different levels of treatment in policy that generated our data
- $\pi_0(z|X_i)$ is our (for now, exactly known) propensity score.

Inverse-probability weighted estimators

- Normalized (Hàjek) estimator

$$\hat{\mu}_{\pi}^{Hàjek} = \frac{1}{\sum_i w_{\pi i}} \sum_{i=1}^{N} Y_i w_{\pi i}.$$

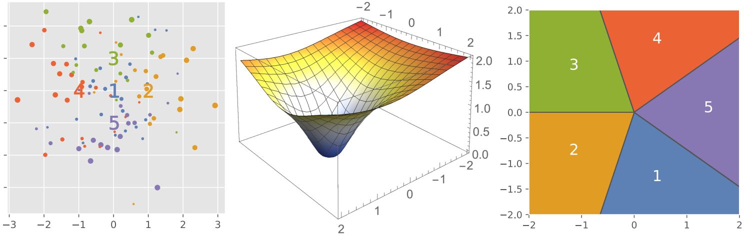
- Note: Previous examples have used implicitly used normalized weights.

Inverse-probability weighted estimators

- Unnormalized (Horvitz–Thompson estimator):
  $\hat{\mu}_{\pi}^{HT} = \frac{1}{N} \sum_i Y_i \frac{\pi(A_i|X_i)}{\pi_0(A_i|X_i)}$
- Horvitz–Thompson estimator is the minimum variance unbiased estimator given our assumptions.
  - Consider N $= 1$. Then $\hat{\mu}_{\pi}^{HT} = Y_i \frac{\pi(A_i|X_i)}{\pi_0(A_i|X_i)}$.
  - ${}_{\pi_0}[\hat{\mu}_{\pi}^{HT}] =_{\pi_0} [Y_i \frac{\pi(A_i|X_i)}{\pi_0(A_i|X_i)}] =_{\pi} [Y_i]$.
- Nonetheless, the normalized (Hàjek) estimator will often have lower MSE.

Policy evaluation

Example with 5 treatments from Kallus (2018).



(a) $X_{1:n}, T_{1:n}$ (b) $\mu_1(x)$ (c) $\pi^*(x)$

Figure: (a) Observed data with 2-dimensional X and observed treatments (color). (b) True mean of treatment 1 as a function of X. (c) Policy we want to evaluate.

Contribution

- Five key features:
  - evaluate any counterfactual polices offline
  - more efficient learning with doubly robust scores
  - use bounded outcome to design a better behavior policy
  - adapt to potential changes in the environment
  - directly optimized on imputed long-term outcome
- Conceptual:
  - bounded outcome encodes information on treatment effect
  - implement a stochastic optimal policy to continue exploration
  - use SI to impute long-term outcome for policy learning
- Practial:
  - increases 1.5-year and 3-year revenue by $15 and $40 per user
  - 3-year total revenue impact sums up to 4-5 millions

Churn prediction

- Train and compare a menu of classification models (logistic regression, random forest, supporter vector machine, xgboost)
- Top performer (xgboost): high overall accuracy (98%), precision (94%), AUC (94%)
- Recall is very low (23%), suggesting that we might be missing some important signals from the constructed features
- Training set:

predicted/actual	0	1
0	35,967	886
1	19	276

- Testing set:

predicted/actual	0	1
0	7731	205
1	4	60

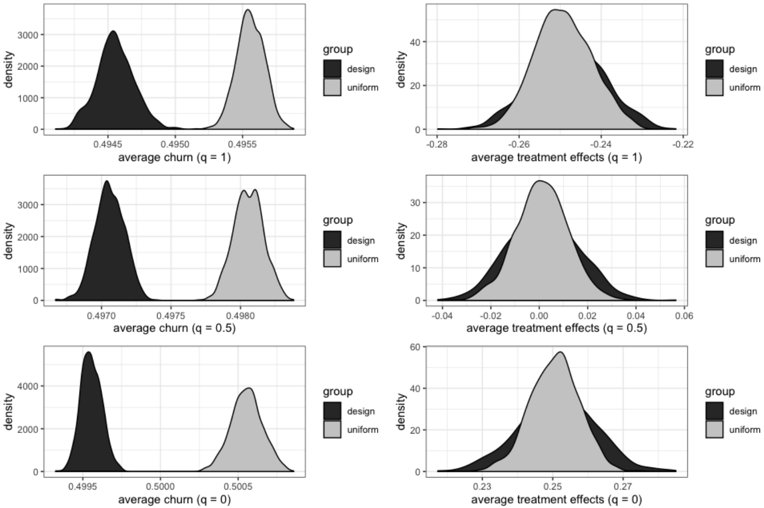
back

Churn prediction

feature	relative importance
credit_card_statusa	100.000
credit_card_statusi	66.728
last_autorenew	39.728
cc_expire_dt	31.951
last_billingchg_reasonremovecc	23.667
first_billingchg_reasonremovecc	18.981
last_start_tenure	7.919
credit_card_typeu	6.016
original_tenure	5.786
last_billingchg	5.252

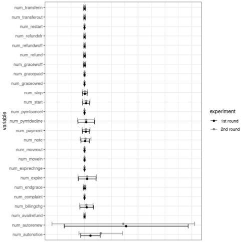
back

Behavior policy simulation

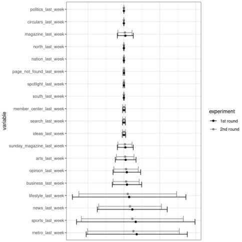


back

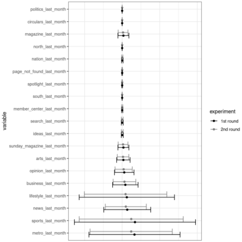
Covariate shift



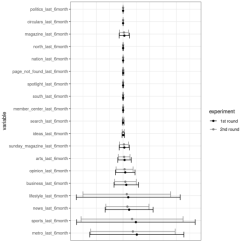
Account Activity



Content Consumption: Last Week



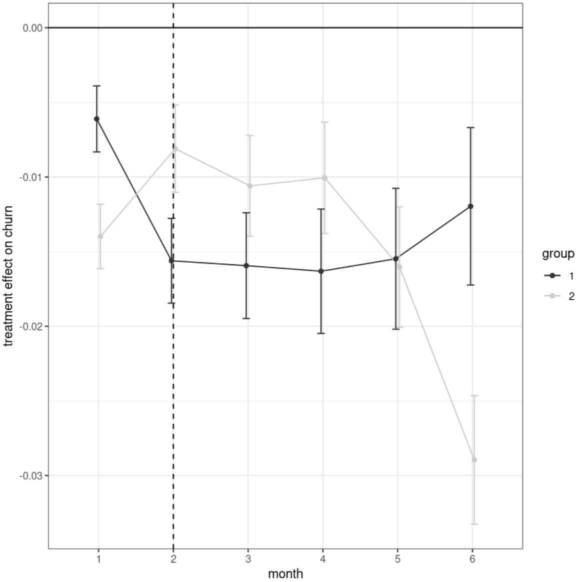
Content Consumption: Last Month



Content Consumption: Last 6 Month

back

Concept shift



back