

Causal Tree に基づく選択バイアスを考慮した条件付き平均処置効果推定手法に関する研究

1X17C089-9 坪井優樹

指導教員 後藤正幸

1. 研究背景・目的

企業がマーケティング施策を講じる際に、適切な効果検証を行い、正しい意思決定につなげることは非常に重要な課題である。観測データから施策効果を正しく評価するために、因果推論という考え方があり、その施策効果推定方法として、一般にランダム化比較実験（以下、RCT）を行うことが望ましいとされている。RCT では、ユーザを施策を行う群（以下、処置群）と施策を行わない群（以下、対照群）にランダムに分割し、効果を推定する。効果を推定する際には、平均処置効果（以下、ATE）と呼ばれる群間の結果の平均値の差や、同じ特徴を持ったユーザ群における ATE である、条件付き平均処置効果（以下、CATE）が用いられる。近年では、どのような特徴を持つユーザ群に対して施策を行うことが有効であるかを分析するために、CATE が多く用いられている。

ここで、CATE を推定する手法として Causal Tree[1] が提案されている。この手法は回帰木のアルゴリズムを用いているため解釈性が高く、施策効果に影響を与える要因についての分析に有用である。しかし、この手法は RCT を前提としているため、選択バイアスに対応できないという問題点がある。ここで、選択バイアスとは、効果が見込まれるユーザ群を人為的に選択して施策を行うような場合に生じる系統的な誤差を意味する。選択バイアスが存在する状況における、ATE を推定する手法として Doubly Robust Estimation（以下、DR 推定法）[2] が提案されている。DR 推定法はユーザの特徴を考慮した ATE 推定手法を 2 つ組み合わせしており、選択バイアスが存在する状況において高い精度で推定が可能であり、加えてノイズに対する頑健性が高い。

そこで本研究では、Causal Tree をベースとし、選択バイアスが存在する状況に対応した CATE 推定手法を提案する。具体的には、DR 推定法を活用した CATE の算出に基づいて、Causal Tree の目的関数の導出を行う方法を示す。これにより、解釈性を維持しつつも選択バイアスが存在する状況において CATE の適切な推定が可能になる。最後に、人工データセットを用いて実験を行い、提案手法の有効性を示す。

2. 準備

2.1. 施策効果の評価

施策効果を表す指標として ATE と CATE がある。ATE は群間の結果の平均値の差であり、集団全体での効果を表す。CATE は共変量と呼ばれる、ユーザの特徴や属性等の因果関係に影響を及ぼすと想定される変数の値が同じ集団における ATE を表す。ここで、ユーザ $i (i = 1, \dots, N)$ の目的変数を $Y_i \in \mathbb{R}$ 、施策の有無を $Z_i \in \{0, 1\}$ とする。ただし、施策を行う場合を $Z_i = 1$ 、行わない場合を $Z_i = 0$ とする。また、ユーザ i における K 次元の共変量を $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{iK})$ とする。 \mathbf{X}_i の値が \mathbf{x} であ

る場合の CATE は式 (1) で定義される。

$$E[Y_i | \mathbf{X}_i = \mathbf{x}, Z_i = 1] - E[Y_i | \mathbf{X}_i = \mathbf{x}, Z_i = 0] \quad (1)$$

2.2. Causal Tree

Causal Tree とは、回帰木を用いた CATE 推定手法であり、一般の回帰木と比較して目的関数が変更されている。Causal Tree は同じ葉ノードに落ちたユーザの共変量の値がなるべく同じになるように枝を分割する。葉ノードの集合を Π 、葉ノードを $l \in \Pi$ とすると、Causal Tree の CATE は式 (2) で表される。

$$\begin{aligned} \tau(\mathbf{X}_i; \Pi) \equiv & E[Y_i | \mathbf{X}_i \in l(\mathbf{x}; \Pi), Z_i = 1] \\ & - E[Y_i | \mathbf{X}_i \in l(\mathbf{x}; \Pi), Z_i = 0] \quad (2) \end{aligned}$$

ただし、 $l(\mathbf{x}; \Pi)$ は $\mathbf{x} \in l$ となる葉ノード $l \in \Pi$ を表す。学習時のデータには、ユーザ i において、 \mathbf{X}_i, Y_i, Z_i が与えられる。Causal Tree では与えられたデータをランダムに 2 分割し、木を作るための訓練データ S^{tr} と作られた木を用いて推定量を出すための推定データ S^{est} に分けて学習を行う。このデータの分割によって過学習を防ぎ、推定精度が向上する。訓練データ数を N^{tr} 、推定データ数を N^{est} 、 $Z_i = 1$ における S^{tr} を S_1^{tr} 、 $Z_i = 0$ における S^{tr} を S_0^{tr} としたとき、学習の目的関数は式 (3) によって表される。

$$\begin{aligned} \widehat{EMSE}_\tau(S^{tr}, N^{est}, \Pi) \equiv & -\frac{1}{N^{tr}} \sum_{i \in S^{tr}} \hat{\tau}^2(\mathbf{X}_i; S^{tr}, \Pi) \\ & + \left(\frac{1}{N^{tr}} + \frac{1}{N^{est}} \right) \cdot \sum_{l \in \Pi} \left(\frac{V_{S_1^{tr}}(l)}{p} + \frac{V_{S_0^{tr}}(l)}{1-p} \right) \quad (3) \end{aligned}$$

ただし、 $V_{S_1^{tr}}(l)$ は葉ノード l に該当する S_1^{tr} の Y_i の分散、 $V_{S_0^{tr}}(l)$ は葉ノード l に該当する S_0^{tr} の Y_i の分散、 p は S^{tr} における S_1^{tr} の割合を表す。この目的関数の最小化は、各葉ノードにおける CATE の二乗和の平均値を最大化しつつ、分散は小さくすることを意味する。

2.3. Doubly Robust Estimation (DR 推定法)

DR 推定法は、選択バイアスが存在する状況において ATE を推定する手法である。ATE 算出時に共変量を考慮することで、選択バイアスを考慮した推定が可能になる。DR 推定法では \mathbf{X}_i と Z_i を説明変数、 Y_i を目的変数とした回帰分析と、共変量が与えられたもとで施策が行われる確率（以下、傾向スコア）を組み合わせて ATE の推定を行う。どちらか一方の手法が正しければ適切な推定を行うことができるため、回帰分析と傾向スコアをそれぞれ用いた場合よりも、ATE の推定精度とノイズに対する頑健性を高めることができる。傾向スコアは式 (4) で表される。

$$e(\mathbf{X}_i) = p(Z_i = 1 | \mathbf{X}_i) \quad (4)$$

$e(\mathbf{X}_i)$ は \mathbf{X}_i を説明変数, Z_i を目的変数とした適当な回帰モデルから事前に推定する. 回帰分析で推定された, $Z_i = 1$ における \hat{Y}_i を $\hat{Y}_{(1)i}$, $Z_i = 0$ における \hat{Y}_i を $\hat{Y}_{(0)i}$ とするとき, DR 推定法における ATE は, 式 (5) で表される.

$$E \left[\frac{Y_i Z_i}{e(\mathbf{X}_i)} - \frac{\hat{Y}_{(1)i}(Z_i - e(\mathbf{X}_i))}{e(\mathbf{X}_i)} \right] - E \left[\frac{Y_i(1 - Z_i)}{(1 - e(\mathbf{X}_i))} + \frac{\hat{Y}_{(0)i}(Z_i - e(\mathbf{X}_i))}{(1 - e(\mathbf{X}_i))} \right] \quad (5)$$

3. 提案手法

3.1. 概要

Causal Tree の CATE $\tau(\mathbf{X}_i; \Pi)$ は, 葉ノード l に該当する群間の結果の平均値の差を表す. そのため, 施策対象者をランダムに選択できない場合, 分岐途中のノードにおいて, 選択バイアスの存在によって正しく推定できない可能性がある. そこで各ノードにおいて, 選択バイアスが存在する状況で正しく ATE を推定できる DR 推定法を活用した CATE の算出に基づいた, Causal Tree の目的関数の導出を行う.

3.2. DR 推定法の活用

DR 推定法を用いた Causal Tree の CATE を式 (6) で表す.

$$\begin{aligned} & \tau_{DR}(\mathbf{X}_i, e(\mathbf{X}_i), \hat{Y}_{(0)i}, \hat{Y}_{(1)i}; \Pi) \\ & \equiv E \left[\frac{Y_i Z_i}{e(\mathbf{X}_i)} - \frac{\hat{Y}_{(1)i}(Z_i - e(\mathbf{X}_i))}{e(\mathbf{X}_i)} \middle| \mathbf{X}_i \in l(\mathbf{x}; \Pi) \right] \\ & - E \left[\frac{Y_i(1 - Z_i)}{(1 - e(\mathbf{X}_i))} + \frac{\hat{Y}_{(0)i}(Z_i - e(\mathbf{X}_i))}{(1 - e(\mathbf{X}_i))} \middle| \mathbf{X}_i \in l(\mathbf{x}; \Pi) \right] \end{aligned} \quad (6)$$

このとき, $V_{S_1^{DR}}(l)$, $V_{S_0^{DR}}(l)$ をそれぞれ式 (7), (8) としたもとの, 提案手法の目的関数を式 (9) で定義する.

$$V_{S_1^{DR}}(l) = V \left[\frac{Y_i Z_i}{e(\mathbf{X}_i)} - \frac{\hat{Y}_{(1)i}(Z_i - e(\mathbf{X}_i))}{e(\mathbf{X}_i)} \right] \quad (7)$$

$$V_{S_0^{DR}}(l) = V \left[\frac{Y_i(1 - Z_i)}{(1 - e(\mathbf{X}_i))} + \frac{\hat{Y}_{(0)i}(Z_i - e(\mathbf{X}_i))}{(1 - e(\mathbf{X}_i))} \right] \quad (8)$$

$$\begin{aligned} & \widehat{EMSE}_{\tau_{DR}}(S^{tr}, N^{est}, \Pi) \\ & \equiv -\frac{1}{N^{tr}} \sum_{i \in S^{tr}} \hat{\tau}_{DR}^2(\mathbf{X}_i, e(\mathbf{X}_i), \hat{Y}_{(0)i}, \hat{Y}_{(1)i}; \Pi) \\ & + \left(\frac{1}{N^{tr}} + \frac{1}{N^{est}} \right) \cdot \sum_{l \in \Pi} \left(\frac{V_{S_1^{DR}}(l)}{p} + \frac{V_{S_0^{DR}}(l)}{1 - p} \right) \end{aligned} \quad (9)$$

4. 評価実験

4.1. 実験条件

提案手法の有効性を示すため, 20 次元の共変量からなる人工データを用いて実験を行う. 学習データ数は 1000 件, テストデータ数は 250 件である. 目的変数 Y_i は 8 個の共変量からなる関数 $b(\mathbf{X}_i)$ と 4 個の共変量からなる関数 $\tau(\mathbf{X}_i)$ を用いて式 (10) に従って生成した.

$$Y_i = b(\mathbf{X}_i) + \frac{1}{2} \cdot (2 \cdot Z_i - 1) \cdot \tau(\mathbf{X}_i) + \sigma \cdot \epsilon_i \quad (10)$$

$b(\mathbf{X}_i)$ は施策の有無 Z_i に無関係な項であり, $\tau(\mathbf{X}_i)$ が施策による効果を表す. ここで, 20 個の共変量 \mathbf{X}_i は各要素が独立な標準正規分布に従うものとし, ノイズ項 ϵ_i も標準正規分布に従うものとした. ただし, ϵ_i の係数 σ によりノイズの大きさを変更し, 本実験では $\sigma = 1, 4, 8, 12$ とした. さらに, 選択バイアスを存在させるため, $\tau(\mathbf{X}_i)$ を構築する 4 個の共変量のうち, 2 個の共変量からなる関数 $e(\mathbf{X}_i)$ で得られるパラメータを持つベルヌーイ分布に従って施策の有無 Z_i を生成する. 評価指標は, 各条件に対して 10 回試行した $\tau(\mathbf{X}_i)$ の RMSE の平均値とする. 比較手法には, 通常の Causal Tree に加えて, 機械学習を用いた CATE 推定手法の 1 つである, X-learner[3] を用いる. DR 推定法と X-learner における傾向スコアはロジスティック回帰により推定し, DR 推定法の回帰分析と X-learner の CATE 推定モデルには XGBoost を使用する. Causal Tree と提案手法の木の深さはそれぞれグリッドサーチにより決定した.

4.2. 結果と考察

各手法の RMSE を図 1 に示す.

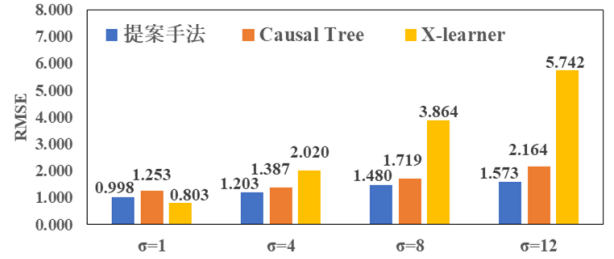


図 1: 各手法の RMSE

全ての条件下で, 提案手法は Causal Tree よりも高い精度を示しており, 選択バイアスが存在する状況において, 有効であるといえる. また, $\sigma = 4, 8, 12$ のノイズが大きい状況において, 提案手法が X-learner よりも高い精度を示した. 以上より, 提案手法は従来手法と比較して高い精度と頑健性を誇り, 選択バイアスとノイズが存在する状況において非常に有効な手法であるといえる.

5. まとめと今後の課題

本研究では, Causal Tree の持つ解釈性を維持したまま選択バイアスが存在する状況に対応することを目的として, DR 推定法を Causal Tree に活用した手法を提案した. また, 人工データを用いた実験により, 提案手法の有効性を示した. 今後の課題として, 処置群と対照群のデータ数が極端に不均衡な状況を考慮することなどが挙げられる.

参考文献

- [1] S. Athey and G. Imbens, "Recursive partitioning for heterogeneous causal effects," *Proceedings of the National Academy of Sciences*, vol.113(27), pp.7353–7360, 2016
- [2] Funk, Michele Jonsson, et al, "Doubly robust estimation of causal effects," *American Journal of Epidemiology*, vol.173(7), pp.761–767, 2011
- [3] Künzel, Sören R., et al, "Metalearners for estimating heterogeneous treatment effects using machine learning," *Proceedings of the National Academy of Sciences*, vol.116(10), pp. 4156–4165, 2019