

研究計画書

2022 年 8 月 12 日

篠崎研究室 学部 4 年 小町雄紀

目 次

1	取り組んだこと	2
2	研究テーマ	2
3	研究背景	2
4	先行研究	2
4.1	TMLE	2
4.2	Super learner	3
5	研究の目的	3
6	研究対象	3
7	現状の課題	3
8	今後の方針	3
9	参考文献	3

1 取り組んだこと

資料をいただいた後に取り組んだこと一覧

- 論文や教科書などで因果推論をはじめとした理論の学習
- [Tlverse](#) を用いて TMLE と Super Learner の理論と R での実装の学習
- [R Guide for TMLE in Medical Research](#) をもとに TMLE の実装 (R+Python)

実装したコードは本計画書の PDF と共に[小町の Github リポジトリ](#)に保存してあります.

2 研究テーマ

現在考えている研究テーマの候補は以下.

- 因果推論手法 (TMLE) を用いて機械学習モデルの構築
- 因果推論手法 (TMLE) を用いてデータ解析

正直なところ, TMLE の学習をした方がいいがそこから何に繋げるか迷っているのが現状.

3 研究背景

TMLE など因果推論手法によって因果効果を推定する取り組みはなされているが, 推定した因果効果を利用した取り組みはなされていない.

4 先行研究

4.1 TMLE

標的型最尤推定量 (target maximum likelihood estimator). アウトカム Y , 暴露 X , 交絡変数 C として推定されるパラメータ

$$\Psi = E[E[Y|X = 1, C] - E[Y|X = 0, C]] \quad (1)$$

は二重ロバスト性を持つ. 二重ロバスト性とはアウトカムに対するモデルと暴露に対するモデルのどちらか一方が正しい場合に因果効果の一致推定量となる推定量である.

傾向スコアの推定や初期の因果効果の推定に機械学習アルゴリズムを適用することができ, 特に次に紹介する Super Learner を使用した TMLE は注目を浴びている. (Laura B. Balzer, Maya L. Petersen, 2021), (Paul N. Zivich, Alexander Breskin, 2020)

近年では改良された TMLE (CV-TMLE, Positivity-C-TMLE) (Cheng Ju, Joshua Schwab, Mark J van der Laan, 2019), (Paul N. Zivich, Alexander Breskin, 2020) も提案されている.

実装の手順は別資料 (TMLE_procedure.ipynb) を参照

4.2 Super learner

機械学習におけるアンサンブル学習の一種, 交差検証の理論をもとに作られており, ある機械学習予測モデルの集合 Ψ の重み付き平均として以下式 1 のように表される. (Eric C. Polley and Mark J. van der Laan,2010)

$$\Psi_{SL} = \sum_{k=1}^n \hat{\alpha} \hat{\Psi}_k \quad (\Psi_k \in \Psi \quad k = 1, 2, \dots, n) \quad (2)$$

実装の手順は別資料 (Super_learner_v1.ipynb) を参照

5 研究の目的

- 因果推論手法 (TMLE) を用いて機械学習モデルの構築

因果推論手法を利用した特徴量の生成, 変数の選択などを提案し, 機械学習モデルの予測精度の向上を図る.

6 研究対象

まだ定まっておらず,8/12 の打ち合わせで決定できればと思います.

7 現状の課題

- 6 の通り, 研究対象が決まっていない
- 因果推論への理解が足りていない

8 今後の方針

- 研究テーマを決定し, できれば中間発表までに使用するデータを決定する.
- 中間発表後は M1 の眞玉さんと機械学習アルゴリズムの勉強会を予定 (週 1)
- 他の院生にもお願いして勉強会をする.

9 参考文献

- Ehsan Karim,Hanna Frank:”R Guide for TMLE in Medical Research”,<https://ehsanx.github.io/TMLEworkshop/>,2021-08-24
- Eric C. Polley,Mark J. van der Laan:”Super Learner In Prediction”,University of California, Berkeley,2010
- Megan S. Schuler,Sherri Rose*:”Targeted Maximum Likelihood Estimation for Causal Inference in Observational Studies”,American Journal of Epidemiology,2016

- Paul N. Zivich, Alexander Breskin: "Machine Learning for Causal Inference: On the Use of Cross-fit Estimators", Department of Epidemiology, 2020
- Tony Blakely, John Lynch, Koen Simons, Rebecca Bentley, Sherri Rose: "Reflection on modern methods: when worlds collide—prediction, machine learning and causal inference", International Journal of Epidemiology, 2020
- K. John McConnell, Stephan Lindner: "Estimating treatment effects with machine learning", Health Services Research,
- Miguel A. Hernán, James M. Robins: "Causal Inference: What If", <https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>, 2020
- Laura B. Balzer, Maya L. Petersen: "Invited Commentary: Machine Learning in Causal Inference—How Do I Love Thee? Let Me Count the Ways", American Journal of Epidemiology, 2021
- Tlverse: <https://tlverse.org/tlverse-handbook/index.html>
- Jenna Wong, Travis Manderson, Michal Abrahamowicz, David L Buckeridge, Robyn Tamblyn: "Can Hyperparameter Tuning Improve the Performance of a Super Learner?", Epidemiology 2019;30: 521–531
- Cheng Ju, Joshua Schwab, Mark J van der Laan: "On adaptive propensity score truncation in causal inference", Statistical Methods in Medical Research, 2019