

ランダムフォレストを用いた系列処置に対する因果的効果の推定

慶應義塾大学大学院理工学研究科 中村知繁

慶應義塾大学理工学部 南 美穂子

1 はじめに

系列処置に対する因果的効果の推定とは、繰り返し複数時点で行われるそれよりも前の時刻で観測された共変量に依存した処置が、結果変数に対してどのように影響しているかを推定することである。系列処置の具体的な例としては、選挙におけるキャンペーンの効果推定 (Blackwell, 2013) や、 β 遮断薬の死亡率への因果的効果の推定 (Williamson and Ravani, 2017)、Antiretroviral Therapy の死亡時刻に対する因果的効果の調査 (Cole et al., 2003) などがあげられる。処置が系列的である場合の因果的効果の推定手法として代表的なもの1つは、Robins(1999) によって提案された Marginal Structural Models(MSM) である。

MSM を用いた系列処置の因果的効果の推定は、まず (i) 処置歴を条件づけたもとの結果変数の期待値に対するモデル (周辺モデル) 及び (ii) 各標本がある処置歴を受ける確率 (傾向スコア) に対するモデルの2つを設定する。その後、(ii) のモデルで推定した傾向スコアの逆数を重み付けた、(i) の推定方程式を解くことで因果的効果を推定する。このとき、Sequential Ignorability Assumption と、Positivity Assumption の2つの仮定 (Blackwell, 2013) のもとで、(i) 及び (ii) のモデルが適切に設定されていれば、得られた因果的効果の推定量は一致性を持つことが知られている。

しかし、実際の解析では、(i) の周辺モデルの設定と、(ii) 傾向スコアに対するモデルを適切に行うことは難しい。実際、傾向スコアは各時刻で処置を受ける確率の積で得られるため、小さな推定の誤差が大きなバイアスを生むことが指摘されている。傾向スコア推定の不安定性から生じるバイアスを減らすための手法として Stabilized Weights (Robins et al., 2000)、や CBPS を用いた傾向スコア推定 (Imai and Ratkovic, 2015) などが提案されている。

2 ランダムフォレストを用いた系列処置の因果的効果の推定

本発表では、系列処置の因果的効果をモデルを設定したもとの推定する MSM とは異なり、Random Forest を用いてノンパラメトリックに推定する手法を提案する。Random Forest は、Breiman(2001) で提案された、Decision Tree または Regression Tree のアンサンブルによって予測子 (predictor) を構成するバギング法の一つである。予測やクラスタリングの問題に対して様々な分野で優れた結果を残している。また、学習を制御するパラメータが少なく、over-fitting しづらいため、利便性が高いことが特徴としてあげられる。

近年、Random Forest によって構成される予測子 (RF 予測子) は、一致性 (Wager and Wether, 2015)、漸近正規性 (Wager and Athey, 2018) を持つことが示され、統計的推論や検定が可能になった。Random Forest の因果的効果推定への応用としては、処置が1回の場合の因果的効果の推定 (Causal Forest, Wager and Athey, 2018)、や操作変数を用いた因果的効果の推定 (Instrumental Forest, Athey et al., 2019) などが提案されている。

本発表では、以上の結果を用いて因果的効果の推定を処置が系列的に行われる場合へと拡張したアルゴリズムを提案し、その理論的な性質について述べた上で、シミュレーションを通して具体例を示す。詳細な結果は当日報告する。

参考文献 (抜粋)

- Blackwell, M. (2013). A Framework for Dynamic Causal Inference in Political Science, *American Journal of Political Science*, 57, 504-520.
- Athey, S., Tibshirani, J. and Wager, S. (2019). Generalized random forest. *Annals of statistics*.
- Robins, J. (1999). Marginal Structural Models Versus Structural Nested Models as Tools for Causal Inference, in *Statistical Models in Epidemiology, the Environment and Clinical Trials*, eds. M. E. Halloran and D. A. Berry, New York: Springer, 95 – 134.
- Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of American Statistical Association*.
- Wager, S. and Walther, G. (2015). Adaptive Concentration of Regression Trees, with Application to Random Forests. arXiv preprint arXiv:1503.06388.