

機械学習を用いた格差と因果効果の推定: Balancing Weights を軸とした手法整理

日時: 2025年2月17日(月) 10:30~17:00

場所: オンライン開催(詳細は別途ご案内)

料金: 一般 5,000円、学生 2,500円

講師: 川田恵介(東京大学)

定員: 35名 ※変更の可能性あり

■本コースの内容

本コースでは、これまで機械学習に触れたことがない学生・研究者を念頭に、数式などを用いない、直感的な議論と R による実装に注力した議論を行います。本年は、社会科学における実証分析の主要テーマである、因果効果や格差の推定への応用に特化し、より多くの実践的な議論の提供を目標とします。特に Double/Debiased Machine Learning または Target Learning と呼ばれる、教師付き学習と伝統的なセミパラメトリック推定の融合領域を紹介し、ます。

格差や因果効果の定量的推定は、社会科学におけるデータ分析の主たる研究関心です。例えば、「性別」間の所得格差の実態を定量的に明らかにする膨大な研究蓄積があります。また出産や結婚などのライフイベントが、働き方や収入などに与える因果効果を推定する研究も、盛んに行われています。

これら二つの研究関心に関連した、多くの推定手法が提案されてきました。例えば伝統的には、重回帰推定や傾向スコアの推定と活用が行われてきました。機械学習(特に教師付き学習)を応用する試みも、近年盛んに行われています。結果、「予測の手法」であると見なされてきた機械学習が、社会科学の伝統的なテーマである格差や因果効果研究と結びつき、機械学習の活用がより盛んになっています。

これらの手法は、R や Python などの代表的な統計言語を用いて容易に実装できます。こうした分析環境整備は、「研究者による根拠なきモデル化」の弊害を緩和し、研究の透明性/信頼性を高めることが期待できます。

しかしながら分析手法の増大は、実践における混乱をもたらす可能性があります。それぞれの手法の特性を理解し、分析課題に応じて選択することが求められます。本コースでは、「Balancing Weight を用いたバランス後の比較」という一般的な推定の枠組みを紹介し、この特殊ケースとして、重回帰や傾向スコア、Double Machine Learning などを位置付けます。さらに「バランス後の比較結果」の事前解釈として、「因果効果の格差の識別」を位置付けます。これにより、より体系的かつ実践的な手法理解を目指します。

コース内容のイメージを掴みたい方は以下のレポジトリをご参照ください。しかし実際の内容とは、異なることをご了承ください。

Github のレポジトリ: <https://github.com/tetokawata/NoteBalance>

■次のような方におすすめです

- ・ 格差研究/因果効果研究/機械学習の社会分析への活用に関心のある方
- ・ 実装方法も同時に学ぶことで、ご自身の研究に迅速に活用されたい方

- ・根拠に基づいた定量的分析を行いたいが、数理的な議論に不慣れな方

■注意事項

- ・どなたでも参加可能です。
- ・ただし、SSJ データアーカイブのデータを利用した講義の場合、3月中に利用したデータを削除して頂く必要があります。
- ・大学または公的研究機関所属の研究者・学生（学部生も可）、SSJ データアーカイブへデータを寄託されている民間研究機関の方は、その後、研究目的で SSJ データアーカイブより申請して頂くことで利用可能です。
- ・R・R studio・必要パッケージのインストールを事前に済ませてください。登録・インストール方法を紹介した動画を事前に配布します

■本コースの日程

- ・バランス後の比較とは？
- ・Balancing Weight とは？
- ・OLS による Balancing Weight の「暗黙の」推定
- ・機械学習のよる Balancing Weight の推定と Double Check
- ・機械学習の個別手法: Random Forest と Stacking
- ・バランス後の比較結果の事前解釈(識別)
- ・バランス後の平均差の推定

*進度によって内容が若干変わることがあります。