

他者への影響と不遵守を考慮した因果推論

—曝露写像と操作変数法によるネットワークデータ分析—

概要

統計的因果推論は、結果に対する原因の影響の大きさを推測するために、いかに統計データを分析できるか考える学問領域です。因果推論の最も標準的なアプローチは、原因となる介入を各被験者に対してランダムに割り当てるランダム化比較実験によって集めたデータを分析することです。従来のアプローチでは、各被験者に対する介入はその被験者のみに影響する、あるいは、すべての被験者は割り当てられた介入を遵守すると想定されることが一般的でした。しかし、実際には、各被験者への介入は複雑な形で他者にも影響すると同時に、割り当てられた介入を遵守しない被験者も存在する状況が数多くあります。

そこで、京都大学大学院経済学研究科 柳貴英 准教授と早稲田大学政治経済学術院 星野匡郎 准教授の共同研究チームは、そのような複雑な波及効果と介入割当の不遵守を考慮できる、曝露写像と操作変数法による新しい因果推論手法を開発し、その統計的性質を証明することに成功しました。

本研究成果は、2023年11月17日に、統計学分野におけるトップジャーナルである「*Journal of the American Statistical Association*」にオープンアクセスでオンライン掲載されました。



1. 背景

ネットワークデータを利用した近年の因果推論では、人々の間に生じる波及効果を要約するためのアプローチとして、曝露写像が考えられています。たとえば、波及効果を推測するために、各人物が介入を受けた友人を持つかどうか注目し、「介入を受けた友人がいる人々にとっての平均的な結果」と「介入を受けた友人がいない人々にとっての平均的な結果」を比べることは、曝露写像アプローチの典型例です。

曝露写像は有用なアプローチであるものの、実際のデータ分析において、曝露写像を適切に選ぶことは容易ではありません。なぜなら、考えられる曝露写像の候補は数多くあることに加えて、波及効果はネットワーク上で複雑に拡がりうるため、正しい曝露写像は分析者が考える候補の中にはないほど複雑な形をしているかもしれないためです（図1）。

このような問題意識をもとに、最新の研究では、誤った曝露写像を選択してしまったとしても因果推論を適切に実施できる方法が開発されています。しかし、これらの研究では、すべての被験者が介入の割当を遵守する理想的なランダム化比較実験のみが想定されてきました。

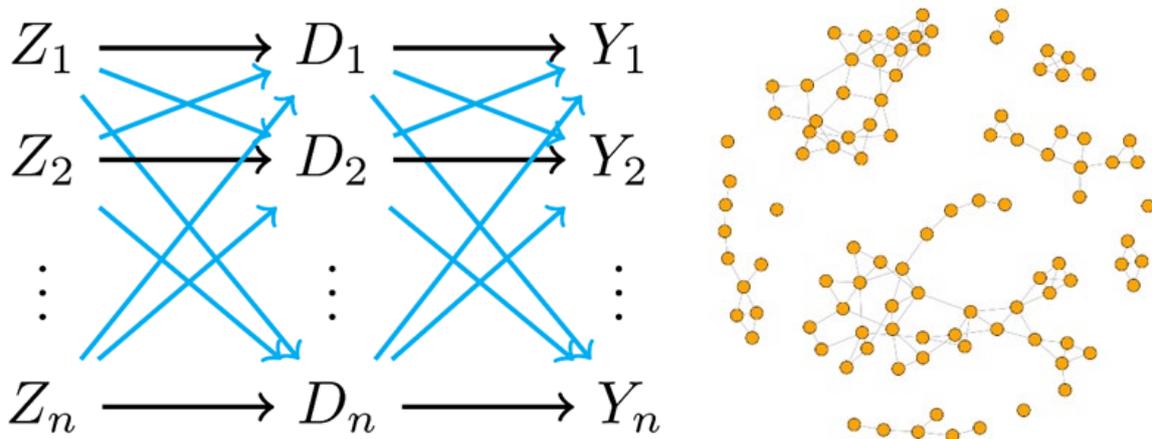


図1 左：複雑な波及効果と介入割当の不遵守がある因果推論の概念図。各変数の右下にある添え字は各人物の識別番号、Zは介入の割当、Dは介入の実現、Yは結果、黒矢印は直接効果、青矢印は波及効果をあらわしています。複雑な波及効果がある場合、たとえば、「友人の友人の友人」からも影響される可能性があります。複雑な波及効果を要約するためのアプローチとして曝露写像が考えられていますが、これまでの研究の多くではZとDが一致しない可能性、つまり、介入割当の不遵守を考慮できていませんでした。

右：波及効果が広がるネットワークのイメージ。オレンジ色の球が人物、球を結ぶ線がデータ収集できる友人関係等の繋がりをあらわしています。波及効果は、ネットワーク上で複雑に広がる可能性があると同時に、考えられる曝露写像の候補は数多くあるため、曝露写像を適切に選ぶことは容易ではありません。

2. 研究手法・成果

本研究では、誤った曝露写像を選択する可能性と介入割当の不遵守の両方を許容したもとの、ネットワークデータを利用して、結果に対する介入の平均的な直接効果・波及効果・全体効果のそれぞれを調べるための新しい因果推論手法を開発しました。

本研究成果は、下記のように構成されます。

1. 因果推論モデルの構築

本研究では、友人関係等の人々の繋がりをソーシャル・ネットワーク・データとして観測できる状

況を想定します。そして、介入割当の不遵守と複雑な波及効果の両方を許すために、割り当てられた介入と実現した介入が異なる可能性、および、各被験者について実現した介入と結果が他者の介入割当や実現した介入に複雑な形で依存する可能性を考慮した因果推論モデルを構築しています。さらに、本研究における曝露写像の正しさとは何かを明確にするために、正しい／誤った曝露写像の定義も与えています。

II. 因果効果パラメータの定義

本研究では、平均的な因果効果をあらわすパラメータを定義するために、遵守者に注目します。遵守者とは、介入を割り当てられたときにのみ介入を受ける人物のことで、操作変数法による因果推論ではお馴染みの概念です。ただし、本研究では、各被験者について実現する介入が自身の介入割当のみならず他者の介入割当にも依存する可能性があるため、これを考慮したもとの遵守者の概念を考えています。

本研究では、誤った曝露写像を選択してしまったとしても、平均的な因果効果として解釈できる、下記のパラメータに注目しています。

- ① **局所平均直接効果**：遵守者に対する介入が自身の結果に与える影響の平均的な大きさ。
- ② **局所平均間接効果**：遵守者に対する介入が他者の結果の合計に与える影響の平均的な大きさ。
- ③ **局所平均全体効果**：遵守者に対する介入が自身と他者の結果の合計に与える影響の平均的な大きさ。

III. 分析方法の開発と統計的性質の解明

本研究では、上記の因果効果パラメータを推測するための方法として、操作変数推定と呼ばれるアプローチを考えています。より具体的には、自身あるいは他者の介入割当を操作変数として採用すると同時に、誤っているかもしれない曝露写像の情報も活用できる、ワルド推定と呼ばれるアプローチを提案しています。さらに、統計的推論のためのアプローチとして、ワルド推定量の近似的な分布にもとづく解析的な方法とネットワーク・ワイルド・ブートストラップと呼ばれる計算機による方法も考案しています。

本研究では、開発した操作変数推定が一致性等の望ましい統計的性質をもつことを証明することで、開発した分析方法を統計理論の観点から正当化しています。そのために、本研究では、下記のような一連の条件が成立する状況を考えています。

- ① **除外制約**：介入割当は実現した介入を通じてのみ結果に影響を与えられる。
- ② **独立性**：介入割当は各被験者に対して独立に割り当てられる。
- ③ **関連性**：割り当てられた介入を遵守する被験者が存在する。
- ④ **単調性**：介入を割り当てられなかったときにのみ介入を受けることを選択する被験者は存在しない。
- ⑤ **制限的な波及効果**：遵守者でない被験者に対する介入割当は自身あるいは他者の結果に影響しない。
- ⑥ **波及効果の漸近的な近傍性**：各被験者に対する波及効果は、ネットワークで近くに位置する人物からの影響で大凡が決まり、ネットワークで遠くに位置する人物からの影響は十分に小さい。
- ⑦ **ネットワーク弱依存性**：各被験者がネットワーク上で直接的に繋がっている人数は少ない。

IV. ソフトウェアの開発

開発手法を簡単に実装できるソフトウェア `latenetwork` を開発しました。本ソフトウェアは、統計分析ソフトウェア R のウェブサイト「CRAN」から入手可能です (<https://cran.r-project.org/package=latenetwork>)。

3. 波及効果、今後の予定

本研究では、介入割当の不遵守を許容したランダム化比較実験によってデータを集められる状況を主に想定しています。ランダム化比較実験が難しい状況において、複雑な波及効果を考慮した因果推論のために、いかに観察データを利用できるかは今後取り組むべき課題のひとつと言えます。

4. 研究プロジェクトについて

本研究は、日本学術振興会科学研究費助成事業 19H01473 (研究分担者：柳貴英)、20K01597 (研究代表者：星野匡郎) 他の支援を受けて実施されました。

<研究者のコメント>

「共同研究者の星野匡郎さんとラフな研究アイデアを話していた研究当初の段階から、私たち自身も面白い研究になるだろうと考えていて、論文が掲載されるまでの間ずっと楽しく研究を進めることができました。本研究成果を統計学分野において伝統ある国際学術誌で発表できたことを大変嬉しく思います。波及効果にかかわる課題はまだまだ残されていると感じていますから、今後も研究を進めていきたいと思っています。」(柳貴英)

<論文タイトルと著者>

タイトル：Causal Inference with Noncompliance and Unknown Interference
(不遵守と未知の干渉を考慮した因果推論)

著者：Tadao Hoshino (Waseda University), Takahide Yanagi (Kyoto University)

掲載誌： *Journal of the American Statistical Association*

DOI： <https://doi.org/10.1080/01621459.2023.2284413>