

Title	博士課程進学者数に関する統計的因果探索と交絡因子の取り扱い
Author(s)	高山, 正行; 小松, 尚登; ファム, テトン; 前田, 高志ニコラス; 三内, 顕義; 小柴, 等; 清水, 昌平
Citation	年次学術大会講演要旨集, 39: 267-272
Issue Date	2024-10-26
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/19523
Rights	本著作物は研究・イノベーション学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Research Policy and Innovation Management.
Description	一般講演要旨

博士課程進学者数に関する統計的因果探索と交絡因子の取り扱い

○ 高山 正行 (NISTEP/滋賀大学), 小松 尚登 (滋賀大学), ファム テトン (滋賀大学/理化学研究所),
前田 高志 ニコラス (NISTEP/学習院大学/滋賀大学/東京電機大学/理化学研究所),
三内 顕義 (NISTEP/京都大学/滋賀大学/東京大学/理化学研究所/国立情報学研究所),
小柴 等 (NISTEP/滋賀大学), 清水 昌平 (NISTEP/滋賀大学/京都大学/理化学研究所)

1 はじめに

我が国の研究力向上の文脈において、博士課程進学等、若手研究者支援に関する政策的重要性は依然として大きく、最近でも関連する調査・研究が盛んに行われている(例えば文献 [吉岡 24])。当グループはこれまで、統計的因果探索アルゴリズム LiNGAM (Linear Non-Gaussian Acyclic Model)[Shimizu06, Shimizu11] を核に、博士課程進学に関する政策的要因に関する因果構造の探索を試み議論してきた [高山 22a, 高山 22b, 高山 22c, 高山 23]。本稿ではそれ等を基にしつつ、未観測共通原因も考慮したものを含む、様々な因果探索手法に基づいて分析を行う。また、それらの結果から、これまで以上に統計的観点から踏み込んだ議論を試みる。

2 研究手法

2.1 使用するデータ

本研究では、国立大学ごとの博士課程進学に関する因果探索の取組 [高山 23] の考え方を改良・アップデートする形で、大学改革支援・学位授与機構 (NIAD) が公開している大学基本情報を中心としてデータを再構築する。大学基本情報は、平成 24 年度から令和 5 年度までの各年度について、国公立大学・短期大学から提供された基礎的な情報を掲載したもので、文部科学省の学校基本調査への回答に基づく。そのため、大学別の修士・博士課程の入学者数や進路別の修了者数等について、変数に組み込むことが可能となる。

特に本研究では、大学基本情報を中心に、以下に示す通り、(1) 博士課程進学に係る変数、(2) 博士課程修了者の修了直後のキャリアに係る変数、(3) 大学の研究環境に係る変数、(4) 博士課程進学直後の経済的支援、に係る変数の 4 つの観点から合計 8 変数を選択し、表 1 の通り、各出典元データを突合してデータセットを構築した。

表 1: 本研究で構築したデータセットの変数と出典元データ

変数の種類	出典元データ
修士課程修了者数 ^{*1}	NIAD 大学基本情報 ^{*2} を 基に加工
博士課程進学者数	
博士課程修了者数	
博士課程修了後の ポストドクター就職者数	
博士課程修了後の 大学教員就職者数	
運営費交付金	各国立大学法人の財務諸表を 基に加工
教員一人当たり学生数	NIAD 国立大学法人の 財務指標等を基に加工
JSPS特別研究員DC1採択者数	JSPSのデータを基に加工

(1) 博士課程進学に係る変数 修士課程修了直後の博士課程進学を対象とし、以下の 2 変数を導入する。

- 修士課程修了者数
- 博士課程進学者数^{*3}

(2) 博士課程修了者の修了直後のキャリアに係る変数 博士課程修了者のアカデミアにおける代表的キャリアとして、ポストドクターもしくは大学教員が想定されるため、以下の 3 変数を採用する。

- 博士課程修了者数
- 博士課程修了直後のポストドクター等就職者数^{*4}
- 博士課程修了直後の大学教員への就職者数^{*4}

(3) 大学の研究環境に係る変数 大学の研究環境に関する指標は様々議論されている。例えば文献 [山下 24] では、

^{*1} 医学部・薬学部の 6 年制カリキュラム修了も含む。

^{*2} <https://portal.niad.ac.jp/ptrt/table.html>

^{*3} あくまで各大学の修士課程修了者のうち、進路として博士課程等へ進学予定とした者を指す。そのため、進学した博士課程の大学が、修士課程の大学と異なる事例も存在することには注意を要する。

^{*4} 博士課程進学者と同様、あくまで各大学の博士課程修了者のうちの各進路の予定者であるため、修了した大学と就職先の大学は必ずしも一致しない。

研究環境の一要素である研究時間の減少について、問題構造と解決方法の例について、大学教員およびマネジメント層へのアンケート結果から構造化している^{*5}。それらを参考に、定量的かつ、国立大学ごとに公開データから取得できるものとして、本研究では、以下の2変数を採用する。

- 運営費交付金の収益額
- 教員一人当たり学生数

なお、大学別の因果探索を試みた先行研究 [高山 23] では、運営費交付金の配分額を変数として採用している。しかし、

- 配分額ベースでは、例えば東海国立大学機構のように、2020年度以降に複数の大学が一つの国立大学法人のもとでのガバナンスに切り替わった場合、大学ごとの配分額まで、文部科学省が決定した金額として確定できないこと
- 過年度の繰り越し分等も含めた収益額の方が、実際の大学ごとの活動をより敏感に反映できる可能性があること
- 令和4年度以降、開示すべきセグメント情報が増え、学部ごとの収益額も財務諸表で公表されるようになったことを利用し、今後、研究分野別の因果メカニズムへの議論にも広げていく可能性があること

に鑑み、本研究では収益額を採用した。また、教員一人当たりの学生数は、毎年度大学ごとに公開されている情報を基に算出可能であり、かつ教員の忙しさを示しうる指標の一つであると解釈し、採用している。

(4) 博士課程進学直後の経済的支援に係る変数 博士課程進学者数に対する原因変数の有力な候補である経済的支援の代表的な例として、本研究では日本学術振興会 (JSPS) の特別研究員制度に着目し、特に、博士課程進学前に採択が定まる、以下の1変数のみを採用する。

- JSPS の特別研究員 DC1 採択者数

以上の考え方により、全86国立大学について、平成24年度から令和4年度までの11か年度にわたって、これらの8変数でデータセットを構築した。1年度・1大学あたり1点、データ点数としては合計 $11 \times 86 = 946$ 点となっている。先行研究 [高山 23] では、令和元年度までのデータしかなかったが、本研究では先述の通り、財務諸表から大学単位での運営費交付金の収益額を採用

^{*5} ただしこの構造化は、文献中でも言及されている通り、“分析者の主観をもとに整理”されている点には、注意を要する。

することでデータ点数を拡大した。

ただし、実際には研究に力を入れている大学もあれば、教育が中心となっている大学もあるなど、本来は大学ごとにそれぞれ強みも環境も異なるため、全ての国立大学を同等に扱い、ひとつの因果グラフで説明しようとすることは適切ではない。そこで後述の通り、本研究では運営費交付金の重点支援類型に基づいた層別化を施した上で、分析を行う。

2.2 因果探索のアルゴリズム

通常の因果探索のアルゴリズムでは、因果グラフの有向非巡回性や、未観測共通原因が存在しないこと等の仮定を置く。一方で本研究では、2.1で述べたデータの性質と、1章で述べた趣旨に鑑み、交絡因子の存在可能性についても、因果グラフを一定の精度で推定できる、Fast Causal Inference (FCI) アルゴリズム [Spirites01a] を採用する。詳細は割愛するが、このアルゴリズムは、ノンパラメトリックな因果探索の主要なアプローチの一つである Peter-Clark (PC) アルゴリズム [Spirites01b] を、未観測共通原因が存在するケースまで適用できるように拡張したものである。また、本研究では FCI での結果の特徴を、未観測共通原因がない仮定での分析と比較する観点から、

- FCI の原型となる制約ベースの PC アルゴリズム
- 構造方程式モデルのような関数形を仮定した DirectLiNGAM [Shimizu06, Shimizu11]

でも因果探索を行い、結果を比較する。以上のアルゴリズムは全て python パッケージとして公開されており、本研究では、PC 及び FCI では “causal-learn” [Zheng24]、DirectLiNGAM では “lingam” [Ikeuchi23] を使用した。

2.3 分析にあたっての条件設定

2.1 と 2.2 に基づいた分析を行う際の条件として、以下の通り設定する。

■ 扱う変数の定義と経時変化を加味する範囲 扱う変数は 2.1 に述べた全ての変数を採用する。これに伴う変数の表記法については表 2 にまとめた。また、遅延時間依存性については、今回は 1 年度分まで入れることとする。つまり、0 年目である時刻 t (年度) とその 1 年後である $t+1$ (年度) の 2 時点を考慮して、時間的因果律も考慮した経時変化を見る。

■ 積の構造的因果モデルの仮定 本研究のような事例では、例えば「修士課程修了者が少ない場合、他の変数を調整しても博士課程進学者はほとんど変化しない」といった、単純な線形モデルでは説明できない構造を有する可能性が高い。むしろ、労働経済学において、線形モデル

表 2: 本研究での分析における各変数の名前と説明。

変数に関する説明	変数名
修士課程修了者数	$x_0(t)$
博士課程進学者数	$x_1(t)$
博士課程修了者数	$x_2(t)$
博士課程修了後の ポストドクター就職者数	$x_3(t)$
博士課程修了後の大学教員就職者数	$x_4(t)$
運営費交付金収益額	$x_5(t)$
教員一人当たり数学生数	$x_6(t)$
DC1採択者数	$x_7(t)$

ではなく指数型モデル（コブ・ダグラス型生産関数）が採用されることも参考に、以下のような指数型の「積の構造的因果モデル」[高山 22a, 高山 22b, 高山 23]を仮定した分析が適切と考えられる。

$$x_i(t) = \prod_j \prod_{\tau=0}^{\tau_{max}} x_j(t-\tau)^{b_{i,j}(t,t-\tau)} e^{e_i(t)} \quad (1)$$

本稿では、構造方程式を仮定しない FCI の分析がメインだが、比較対象とする DirectLiNGAM での分析で、特に先行研究で「積の構造的因果モデル」が用いられている。分析するデータの条件を揃えての比較のため、データセットの各値について*6、全て自然対数をとった上で、各アルゴリズムでの因果探索を行うこととする。

■国立大学の重点支援 3 類型に応じた層別化 先述の通り、国立大学の性質に応じた層別化にあたっては、文部科学省による運営費交付金の重点支援にあたっての 3 類型を、本研究では採用する。各類型の概要は主に以下の通りである。

- 重点支援類型 I: 主として、地域に貢献する取組とともに、専門分野の特性に配慮しつつ、強み・特色のある分野で世界・全国的な教育研究を推進する取組を中核とする国立大学（全 55 大学）。
- 重点支援類型 II: 主として、専門分野の特性に配慮しつつ、強み・特色のある分野で地域というより世界全国的な教育研究を推進する取組を中核とする国立大学（東京医科歯科大学、政策研究大学院大学等、全 15 大学）。
- 重点支援類型 III: 主として、卓越した成果を創出し

ている海外大学と伍して、全学的に卓越した教育研究、社会実装を推進する取組を中核とする国立大学（東北大学、京都大学等、全 16 大学）。

ただし、特に重点支援類型 II・III については、重点支援類型 I に比べて大学数が少なく、データ点数も減るため、分析上の精度にも何らかの影響する可能性がある。この影響を軽減するための工夫として、本研究では、層別化された各サブセットから、それぞれ 1,000 回の同じデータ点数の無作為復元抽出を行い、FCI で因果グラフを出力する、bootstrap 法の工程を繰り返し、各辺の出現については bootstrap 確率で定量的に評価することとした。

3 結果・考察

3.1 FCI で見られた因果グラフの全体像

図 1 には、bootstrap 法に基づいて FCI を計算し、各辺の出現確率を因果グラフに付記する形で示している。図 1(a)-1(c) を比較すると、まず重点支援類型 II の場合のみ、

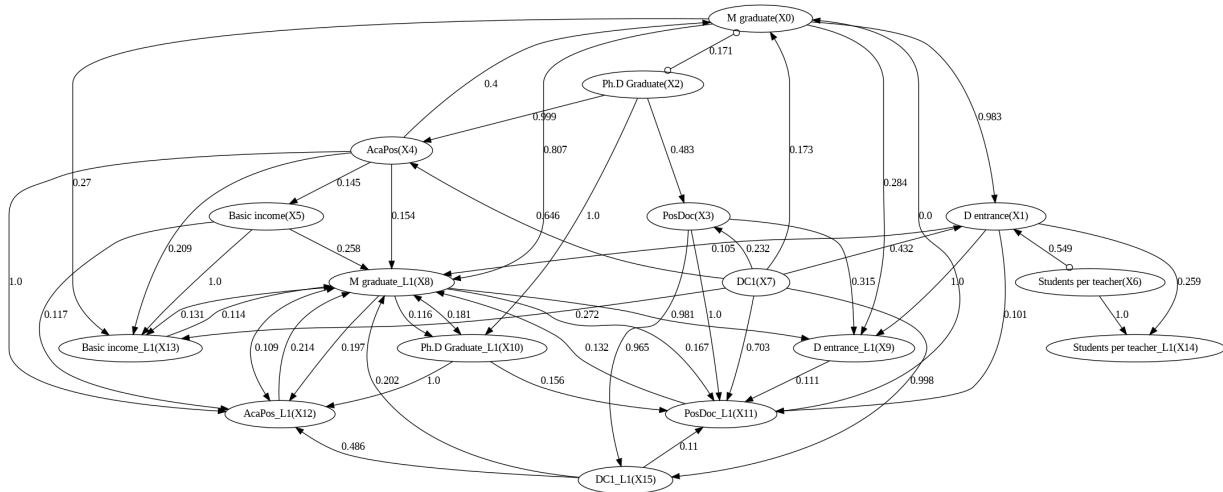
- “ $\circ \rightarrow$ ”、“ \leftrightarrow ”、“ $\circ - \circ$ ”のような、2 変数間の因果関係が決定できないような辺が現れていない
- 「修士課程進学者数」→「博士課程進学者数」などの、全体と内数の関係になっている変数間に表れる辺や、自己回帰など、解釈はともかく、統計的性質を踏まえると自然な辺が中心となって表れており、それ以外の辺があまり見受けられない

といった特徴が見受けられる。これは、重点支援類型 II の大学数がそもそも少ないことや、医学などの特定分野に特化した大学、および大学院大学等、特殊な大学で構成されているグループであることが原因だと考えられる。そのため、以降は特に重点支援類型 I/III の間で比較しながら、特徴的な辺について考察を行う。

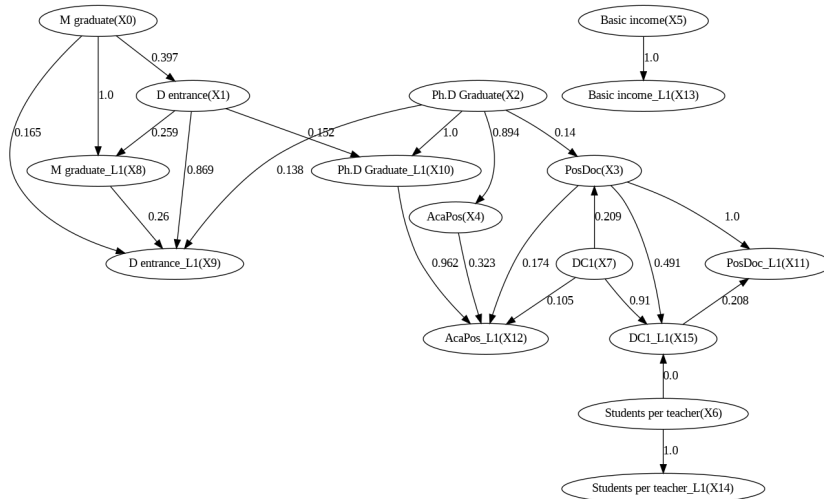
■直接の因果関係が示唆される辺 まず、図 1(a), 図 1(c) の両方で見られる、「DC1」→「博士課程進学者数」という辺は、経済的支援を含む DC1 の採択数の変動が、博士課程進学者数の変動に繋がるという、従来からの政策的議論でも見られた、自然で解釈しやすい関係*7であり、領域知識から見ても妥当な辺である。一方、この bootstrap 確率は、類型 I では 0.432 となっているのに対し、類型

*6 ただしゼロの場合には、対数値の発散を防ぐため、値を 1.0×10^{-5} に置き換えて、因果探索を実行している。

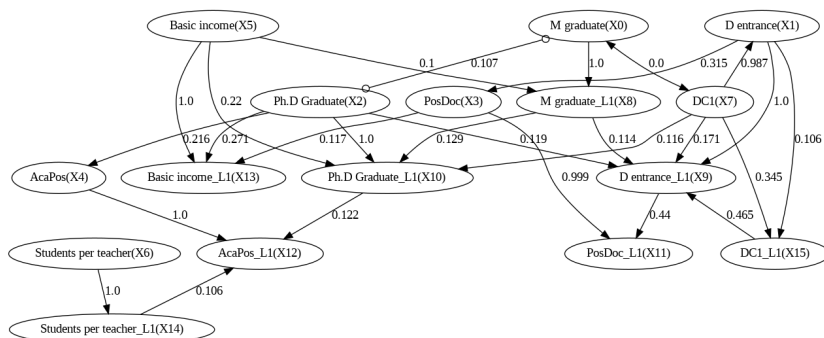
*7 ただし、PC や FCI といった、関数系の仮定もないノンパラメトリックな因果探索では、あくまで条件付き独立性に基づいた探索結果であり、相関の正負によらない結果であることから、この因果グラフをもって経済的支援を増やせば博士課程進学者数の増加につながるものと解釈できるわけではないことに注意が必要である。



(a) 重点支援類型Ⅰの55大学に対するFCIの結果。



(b) 重点支援類型Ⅱの15大学に対するFCIの結果。



(c) 重点支援類型Ⅲの16大学に対するFCIの結果。

図 1: 各重点支援類型ごとの FCI の計算結果。なお、各辺に付した値は、1,000 回の無作為復元抽出に基づき計算された、その辺が現れる bootstrap 確率を示す。また、各辺は以下のように定義される。

$A \rightarrow B$: A が B の原因である。ただし、別途両者に未観測共通原因が存在する可能性はある。

$A \circ \rightarrow B$: B が A の原因となることはない。また、『A が B の原因となる』、あるいは『A と B の間に未観測共通原因が存在する』の少なくとも一方が成り立つ。

$A \leftrightarrow B$: A と B は互いに相手の原因にはならないが、未観測共通原因がある。

$A \circ \circ B$: 『A が B の原因』、『B が A の原因』、『A と B の間に未観測共通原因が存在する』のうち少なくとも 1 つが成り立つ。

Ⅲでは0.987と極めて高い数字となっている。この結果から、類型Ⅰのような多くの大学でも、DC1のような経済的支援が博士課程進学者数に影響するものの、その確からしさは、採択者数が多い研究大学で構成される類型Ⅲにおいて顕著である。

他には、類型Ⅰでは博士修了直後のポスドク就職者数の変化が、翌年のDC1の採択者数に変動に繋がるというbootstrap確率が0.965と高い*8一方、類型Ⅲでは博士課程進学者数の変化が翌年のDC1採択者数の変動に繋がるbootstrap確率が0.106となっている。この傾向は、先行研究[高山23]でも類似の傾向が議論されており、DC1採択者数の増減の要因が、大学の性質によって異なることを示唆しているものと考えられる。

■未観測共通原因の存在が示唆される辺 本研究でFCIを採用して因果探索を行ったことで、これまでデータ駆動の観点からの議論が難しかった、博士課程進学に関する因果グラフの構築における未観測共通原因の位置が、管見の限りでは初めて探索・議論されることになる。具体的には、類型ⅠとⅢでの共通の振る舞いとして、修士課程修了者数と博士課程修了者数は“o-o”で結ばれるbootstrap確率が一定程度ある。つまりこの2変数には、何らかの未観測共通原因を介した相関の可能性がある。特に、類型Ⅰでは1年後(L1)の修士課程修了者数と博士課程修了者数が“↔”で結ばれるbootstrap確率が0.181だけあることから、この2変数間で“o-o”となっている部分の正体は、実際には“↔”の関係、つまり未観測共通原因があるという可能性もある。ナイーブには、両変数はともに修士・博士の学位取得者数という観点から、例えば学位取得のための成果の生産性に関連する変数であること等が想像される。

3.2 他の因果探索アルゴリズム(PC/DirectLiNGAM)との結果の比較

本稿では、紙面の制限があることから簡単な説明のみとするが*9、PCやDirectLiNGAMで同様に因果探索した結果は以下の通りである。

■FCIで直接の因果関係として示唆された辺の妥当性 例えばFCI、PCでデータセット全体で因果探索をした結果は、そのアルゴリズムの類似性による部分もあるが、未観測共通原因の存在が示唆される辺以外は、ほとんど同じ因果構造が出力されることが確認された。ま

*8 実は類型Ⅱでも、類型Ⅰほど高いbootstrap確率ではないが同様の傾向が見えている

*9 これらも含め、本稿の研究の詳細は、<https://github.com/mas-takayama/Causal-Discovery-in-Doctoral-Problem-in-Japan>で公開予定である。

た、DirectLiNGAMでも、3.1節で議論したような、類型Ⅰで博士課程修了直後のポスドク就職者数、類型Ⅲで博士課程進学者数が、翌年のDC1採択者数に影響するという辺、そしてDC1採択者数の変化の同じ年度の博士課程進学者数への影響を示す辺の出現が、概ね確認されている。ここまで議論してきた有向辺については、未観測共通原因を考慮しない場合でも、仮定の異なる複数のモデルで、さらに未観測共通原因を考慮した場合でも頑健に表れていることから、これらは統計的観点からは非常に確度の高い因果関係であると期待される。

■各辺の因果的影響の正負 上記各辺については、DirectLiNGAMで探索された場合に出力された因果係数がいずれも正の値であった。このことから、DC1のような経済的支援の増加は、博士課程進学者数の増加につながることを期待されるという、政策研究の領域知識からも自然に解釈されるような因果関係が、統計的にも示唆された。また、類型Ⅰでは博士課程を終えてそのままポスドクになったロールモデルが増えることで、後輩のDC1採択者数を増やすことにもつながっている一方、類型Ⅲでは直近で博士課程に進学した先輩学生が多いことで、1学年下の修士学生のDC1採択の機会増につながっている等といった、異なるメカニズムが働いているという仮説にもつながり得る*10。

4 おわりに

4.1 本稿での成果のまとめ

本稿では、国立大学での博士課程進学等に関するデータセットについて、先行研究[高山23]以上にデータ点数が多くなるような見直しを行い、統計的因果探索アルゴリズムであるFCIを中心に用いた分析を行った。その結果、複数のモデルで、DC1の採択者数に係る因果関係の可能性が統計的に提示され、さらにFCIからは未観測共通原因の存在位置も示唆された。

4.2 本研究の限界と今後の課題

■政策研究の領域知識に基づく因果グラフの各辺の詳細な解釈 本研究で得られている因果探索結果の大部分の辺については、向きの妥当性等も含めた網羅的な検証・議論は、本稿では行っていない。一方、この領域知識から見た妥当性と統計的妥当性の両方をクリアした因果モデルの構築と理解には、かなりのコストを要する。近年は大規模言語モデルと統計的因果探索を組み合わせ、領

*10 ただし、これらの仮説を正当化するには、アンケート調査等を精緻に分析するなどの取組も必要であり、あくまで統計的因果探索において、統計的信頼性が一定担保された辺を解釈するにあたっての仮説であることに、留意されたい。

域知識から見ても妥当性を担保し、かつ統計的にも妥当性の高い因果モデルを構築する方法等も提案されている [Takayama24]。本取組も、そのような手法を政策研究の文脈で適切に援用することで、効率的かつ正確に領域知識からの妥当性を担保した因果グラフが構築できることが期待される。

■未観測要因の具体的な特定 FCI などの因果探索手法で、因果グラフ上の交絡因子の位置候補が示されても、具体的にその交絡因子がどういふ変数なのかは、特にデータが足りない場合には、領域知識から推測し、何らか適切な補完を検討するしかない。最近では、構造化データ（テーブル形式のデータ）ではなくテキストデータのような非構造化データから、因果関係を捉えた低次元の特徴表現や潜在変数を学習する、いわゆる因果表現学習の方法についても様々検討されており、将来的にはこの手法を用いて、変数を特定していくなどの方法も考えられる。

■研究分野別の議論 実際の博士課程進学や研究環境等の問題は、例えば人文・社会科学や理工学、医学等の各分野によっても状況が異なることから、因果グラフの構造やメカニズムそのものが異なる点も、既に指摘されている [高山 22a]。一方、令和 4 年度から開始した国立大学の第 4 期中期目標期間より、各大学の財務諸表において、セグメント別の情報開示が、学部単位にも広げられるようになったことから、そのような分析も可能になりつつある。本稿では分野別の議論にはこれ以上立ち入らないが、既に大学別・分野別の分析も試みており、講演では時間が許せば紹介する。

■因果モデルを用いたシミュレーション 因果モデルの構築とそのメカニズムの解明は、政策科学という観点で重要なだけでなく、将来的な政策提言への貢献が要請される。本研究の時点で、博士課程進学の問題については、実際の統計データに基づいた因果モデルの構築の目的が一定程度立ったといえる。今後、より具体的な政策意思決定支援を念頭に、“analysis” から “synthesis” に踏み込んで、マルチエージェントシミュレーション等を用いた複数シナリオを構築するといった議論にも応用・発展することが期待される。

謝辞

本研究の一部は、JST、CREST、JPMJCR22D2 の支援を受けて実施したものである。

参考文献

- [高山 22a] 高山正行, 小柴等, 前田高志ニコラス, 三内顕義, 清水昌平, 星野利彦: 博士課程進学率に関する因果モデルの構築: 統計的因果探索アルゴリズム “LiNGAM” による試行的分析. *Jxiv*, 2022. (preprint) <https://doi.org/10.51094/jxiv.1>
- [高山 22b] 高山正行, 小柴等, 前田高志ニコラス, 三内顕義, 清水昌平, 星野利彦: 統計的因果探索アルゴリズム “LiNGAM” を活用した専攻分野別の博士課程進学に関する研究. *研究イノベーション学会 第 37 回年次学術大会 (予稿集)*, 公演番号 1C06, 2022.
- [高山 22c] 高山正行, 小柴等, 川村真理: ペイジアンネットワークによる博士課程進学要因の試行的分析. *研究イノベーション学会 第 37 回年次学術大会 (予稿集)*, 公演番号 1C07, 2022.
- [高山 23] 高山正行, 小松尚登, ファム テトン, 前田高志ニコラス, 三内顕義, 小柴等, 清水昌平: 大学別の博士課程進学等に関するデータセットの構築と統計的因果探索. *研究イノベーション学会 第 38 回年次学術大会 (予稿集)*, 公演番号 2D20, 2023.
- [山下 24] 山下 泉, 西川 開, 村上 昭義, 伊神 正貫 科学技術の状況に係る総合的意識調査 (NISTEP 定点調査 2023) 報告書. *NISTEP REPORT*, No.201, 2024. <https://doi.org/10.15108/nr201>
- [吉岡 24] 吉岡 (小林) 徹, 片岡純也, 横田一貫, 柴山創太郎, 川村真理: 博士課程修了者の職務満足度が高まる要因についての探索的研究. *文部科学省 科学技術・学術政策研究所 Discussion Paper*, No.233, 2024. <https://doi.org/10.15108/dp233>
- [Ikeuchi23] Takashi Ikeuchi, Mayumi Ide, Yan Zeng, Takashi Nicholas Maeda, and Shohei Shimizu: Python package for causal discovery based on LiNGAM. *Journal of Machine Learning Research*, 24(14):1–8, 2023. <http://jmlr.org/papers/v24/21-0321.html>
- [Kadowaki13] K. Kadowaki, S. Shimizu, and T. Washio: Estimation of causal structures in longitudinal data using non-Gaussianity. *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, Southampton, UK, pp.1-6, 2013. <https://doi.org/10.1109/MLSP.2013.6661912>
- [Shimizu06] Shohei Shimizu, Patrik O. Hoyer, Aapo Hyvärinen, and Antti Kerminen: A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7:2003-2030, 2006. <https://www.cs.helsinki.fi/group/neuroinf/lingam/JMLR06.pdf>
- [Shimizu11] S. Shimizu, T. Inazumi, Y. Sogawa, A. Hyvärinen, Y. Kawahara, T. Washio, P. O. Hoyer and K. Bollen.: DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model. *Journal of Machine Learning Research*, 12(Apr): 1225–1248, 2011. <https://dl.acm.org/doi/10.5555/1953048.2021040>
- [Shimizu12] Shohei Shimizu: Joint estimation of linear non-Gaussian acyclic models. *Neurocomputing*, 81:104-107, 2012. <https://doi.org/10.1016/j.neucom.2011.11.005>
- [Spirtes01a] Peter Spirtes: An Anytime Algorithm for Causal Inference. *Proceedings of Machine Learning Research*, R3:278–285, 2001. <https://proceedings.mlr.press/r3/spirtes01a.html>
- [Spirtes01b] P. Spirtes, C. N. Glymour, R. Scheines, D. Heckerman: Causation, prediction, and search(2nd Edition). *MIT press*, 2001. <https://doi.org/10.7551/mitpress/1754.001.0001>
- [Takayama24] Masayuki Takayama and Tadahisa Okuda and Thong Pham and Tatsuyoshi Ikenoue and Shingo Fukuma and Shohei Shimizu and Akiyoshi Sannai: Integrating Large Language Models in Causal Discovery: A Statistical Causal Approach. *arXiv*, 2024. (preprint) <https://arxiv.org/abs/2402.01454>
- [Zheng24] Y. Zheng, B. Huang, W. Chen, J. Ramsey, M. Gong, R. Cai, S. Shimizu, P. Spirtes, K. Zhang: Causal-learn: Causal discovery in Python. *Journal of Machine Learning Research*, 25(60):1–8, 2024. <https://www.jmlr.org/papers/v25/23-0970.html>