

| | |
|--------------|---|
| Title | ベ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：「修士課程追跡調査」を用いた分析の試行 |
| Author(s) | 高山，正行；小柴，等；川村，真理 |
| Citation | 年次学術大会講演要旨集，37：198-203 |
| Issue Date | 2022-10-29 |
| Type | Conference Paper |
| Text version | publisher |
| URL | http://hdl.handle.net/10119/18678 |
| 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 |

ベ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： 「修士課程追跡調査」を用いた分析の試行

○ 高山 正行 (NISTEP, MEXT), 小柴 等 (NISTEP), 川村 真理 (NISTEP, 東京大学)

1 はじめに

我が国の科学技術・イノベーション政策の中核的な課題のひとつである若手研究者の支援は、そのフェーズに応じて政策的議論が異なる。中でも、研究者の主たる源泉である博士人材については、特にその層の厚みを維持・強化していくため、修士課程学生の博士課程進学率向上が重要とされている。他方で、博士課程進学に関わる各種要因及び要因間の因果関係は必ずしも明らかになっておらず、よりの確な政策立案の支援や EBPM の観点からも、データに基づく詳細な因果モデルの構築と、それに向けた様々な数理モデルの活用が期待されている。

こうした背景の元、統計的因果探索を軸とした新たなアプローチも提案されており [高山 21a]、その実践として、博士課程進学率や経済的支援、研究環境、キャリアパスに関する記述統計値からなるデータセットを構築、統計的因果探索アルゴリズム “DirectLiNGAM” を用いて因果グラフ導出を試みる研究も行われている [高山 21b, 高山 22a, 高山 22b]。ただし、これらの試行的分析においては課題も指摘されており、データ点数の充実化による統計的信頼性の向上や、時系列に伴う遅延効果の扱い、選定する変数系の妥当性と事前知識の充実化が挙げられている [高山 22a, 高山 22b]。これらの課題のうち「データ点数の充実」については、NISTEP の「修士課程（6 年制学科を含む）在籍者を起点とした追跡調査」 [治部 21]（以下、「修士課程調査」と呼ぶ。）をはじめとする様々な博士課程進学に関する調査の個票データ活用による解決が期待されている [高山 22a, 高山 22b]。

修士課程調査は、修士課程在籍者に対して、その課程における研究活動の状況や進路決定に関して重視する点まで、様々な観点から博士課程進学に関する意識調査を行ったものであり、博士課程進学率を向上させるための経済的支援や博士課程修了後のキャリアパスの充実の重要性が示唆されている [治部 21]。サンプルサイズも全体で 10,000 件を超えており、これらの個票データを活用した因果探索により、統計的にも尤もらしい、本政策領域での因果に関する新たな知識の獲得が期待できる。

そこで本研究では、若手研究者支援の政策領域における新たな領域知識の獲得を目的として、修士課程調査のデータを対象に、それらデータの特徴を考慮して、因果探索手法の中でも特にカテゴリ変数・離散データと親和性の高い（因果）ベイジアンネットワーク (Bayesian Network; BN) を活用し、因果グラフの導出を試行する。

2 ベイジアンネットワークの活用

■統計的因果探索とベイジアンネットワーク 因果推論、特に統計的因果推論とはデータから因果効果を統計的に推定する手法である。この際、目的変数とそれに関連する要素を集め、背景知識等を元に因果関係を設定し、データからその妥当性を定量的に検証する形で因果効果を推定することが多い。代表的な手法には、回帰分析や操作変数法、差分の差分法、共分散構造分析などが挙げられる。

因果推論の中でも、因果関係自体をデータから自動的に見つけ出す（推測する）ような取組が因果探索に相当する。何らかの背景知識を前提とするその他の手法と比較して、背景知識の必要性が低く、より多くの要素間の複雑な関係性も考慮できるメリットがある。因果探索の代表的な手法としては BN, LiNGAM [Shimizu06] が挙げられる。これらの内容について図 1 にまとめた。

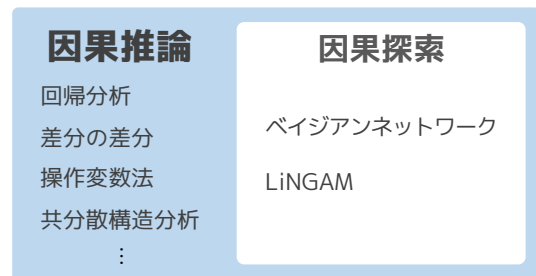


図 1: 因果推論と因果探索の関係

BN も LiNGAM も、どちらも統計的因果探索の手法として尤度の高い因果グラフをデータ駆動的に導くものであるが、その原理の違いから得意とするデータの性質や因果グラフの導出方法は大きく異なる。

因果グラフの種類としては、BN も LiNGAM も原則として有向非巡回なグラフ (Directed Acyclic Graph; DAG) を前提とする。その上で BN は条件付き確率分布 (ベイズ的な枠組み) に基づいて因果関係を表現する [Pearl 85] *1。一方、LiNGAM は独立成分分析をベースに因果関係を導き出す [Shimizu06, Shimizu11, 清水 17]。そのため表 1 に示したとおり、変数の種類として BN は離散値、LiNGAM は連続値のデータを得意とする。

表 1: 因果探索手法 BN と LiNGAM の比較

| 手法 | 変数の種別 | アプローチ |
|--------|-------|--------|
| BN | 主に離散 | ベイズ |
| LiNGAM | 主に連続 | 独立成分分析 |

*1 BN: Bayesian Network

因果探索手法としては BN の方が歴史が長く、様々な分野での利用例があるが、例えば政策研究の文脈における事例としては、鳥海らによる地域健康政策における利用例 [鳥海 18] が挙げられる。本研究でも鳥海らの事例と同様にアンケート調査のデータを分析対象としており、変数の多くが離散値をとるため分析手法としては BN の方が適している。

■本研究で用いる BN の構造学習アルゴリズムの概要
BN の構造学習アルゴリズムには大きく分けて 3 タイプが存在する。以下にその概略を述べる。

スコアリングによる構造学習 観測データに対するネットワークモデルの適合度を示すスコア (赤池情報量基準, ベイズ情報量基準等) を最適化するような DAG を探索し出力する。代表的なアルゴリズムとして山登り法 (Hill-Climbing search; HC) による探索やタブー探索 (Tabu search) などが挙げられる。**条件付き独立性検定による構造学習** 独立性検定・条件付き独立性の検定に基づいて、変数間の関係性について枝刈りを行い、因果グラフのスケルトン構造を構築し、オリエンテーションルールに基づいて部分的に方向づけがなされた DAG (Partially DAG; PDAG) を描写するというサイクルを繰り返すことで、DAG を推定する。代表的なアルゴリズムとして、PC (Peter Spirtes and Clark Glymour) アルゴリズム、GS (Grow-Shrink) アルゴリズムが挙げられる。

*1 なお、BN そのものは因果探索以外に、仮定した DAG への当てはまり度合いを評価した因果推論に使われることもあるが、最適な DAG を導き出す構造学習アルゴリズムも作成されており、因果探索における BN は専ら、この構造学習のことを指す。

ハイブリッド型構造学習 上記 2 つのアプローチをハイブリッドしたもの。代表的なアルゴリズムには、MMHC (Max-Min Hill-Climbing) アルゴリズムなどが挙げられる。

これらアプローチと代表的なアルゴリズムの対応を、表 2 にまとめた。特に HC, PC や MMHC は Python や R における BN 関連のパッケージ・ライブラリでも実装されており利用しやすい。

ところで、BN と言ってもアプローチ・アルゴリズムごとに原理や最適化指標が異なることから、同じデータを対象としても同じ結果が得られるとは限らない。そこで本研究では、まずはデータそのものが持つ各変数の条件付き独立性の評価が信頼性の高いものであるという立場に立ち、複数のアルゴリズムで因果探索を実行して計算結果の大域的な特徴比較を行うこととした。

表 2: BN における主な構造学習の種類とアルゴリズム

| 構造学習の種別 | 主なアルゴリズム |
|------------------|---|
| スコアリングによる構造学習 | ・ HC アルゴリズム ・ Tabu アルゴリズム |
| 条件付き独立性検定による構造学習 | ・ PC アルゴリズム ・ GS アルゴリズム |
| ハイブリッド型構造学習 | ・ H2PC アルゴリズム ・ MMHC アルゴリズム ・ IAMB アルゴリズム |

3 分析対象とデータ

■「修士課程調査」の特徴 分析対象である修士課程調査 [治部 21] の概要と特徴は以下のとおりである。

この調査は 2020 年度の修士課程 (6 年制学科を含む) 修了 (卒業) 者及び修了 (卒業) 予定者を対象とし、

- ・ 在籍中における経済的支援状況
- ・ 進路状況
- ・ 博士課程に進学しない理由
- ・ 在籍者の視点から博士課程進学率を向上させるための政策

等について調べたもので、2020 年 11 月 16 日から 2020 年 12 月 28 日にかけて悉皆で実施され、有効回答数 16,311 件となっている*2。上述の項目だけでなく、専攻分野分類、研究活動の状況、修士課程における満足度やインターンシップの経験の有無等についても調査を行っ

*2 対象者数 125,418 名 (学校基本調査)、回答率: 13.0%, 有効回答率: 13.0%

ているため、本個票データの活用により、在籍者の属性や、経済的支援や研究環境についての様々な状況、博士課程進学に関する考え方と実際の進路の関係について、多面的な分析が期待できる。

表 3: 本研究において個票データから抽出した質問と回答類型

| 質問番号 | 内容 | 回答類型 |
|--------|---|--|
| Q12 | 分野分類 | 01:理学, 02:工学, 03:農学, 04:医学, 05:保健, 06:人文, 07:社会, 08:家政, 09:教育, 10: 芸術・その他 |
| Q16 | 社会人経験 | 1: 有, 2: 無 |
| Q21 | 授業料免除 | 1: 有, 2: 無 |
| Q22 | 借入金 | 1: 有, 2: 無 |
| Q23 | 修了後の進路 | 1: 博士への進学またはその準備, 2: その他 |
| Q35-01 | (博士進学に必要なと考える要件) 進学・編入が容易になる | 1: そう思う, 2: そう思わない |
| Q35-02 | (博士進学に必要なと考える要件) 経済的支援拡充 | 1: そう思う, 2: そう思わない |
| Q35-03 | (博士進学に必要なと考える要件) 優秀な学生が集まる | 1: そう思う, 2: そう思わない |
| Q35-04 | (博士進学に必要なと考える要件) 研究に関与しない学内事務の削減 | 1: そう思う, 2: そう思わない |
| Q35-05 | (博士進学に必要なと考える要件) 実験設備等の研究環境の充実 | 1: そう思う, 2: そう思わない |
| Q35-06 | (博士進学に必要なと考える要件) 時に「侮辱的」「抑圧的」 「敵対的」な文化的環境改善 | 1: そう思う, 2: そう思わない |
| Q35-07 | (博士進学に必要なと考える要件) PI の指導力向上 | 1: そう思う, 2: そう思わない |
| Q35-08 | (博士進学に必要なと考える要件) インターンシップや共同研究で企業 と繋がる | 1: そう思う, 2: そう思わない |
| Q35-09 | (博士進学に必要なと考える要件) 国際学会・留学等の機会増加 | 1: そう思う, 2: そう思わない |
| Q35-10 | (博士進学に必要なと考える要件) 産業界で幅広く活躍するためのスキ ルの習得 | 1: そう思う, 2: そう思わない |
| Q35-11 | (博士進学に必要なと考える要件) アカポスへの就職可能性増加 | 1: そう思う, 2: そう思わない |
| Q35-12 | (博士進学に必要なと考える要件) 若手ポストの雇用条件改善 | 1: そう思う, 2: そう思わない |
| Q35-13 | (博士進学に必要なと考える要件) 民間での博士の雇用増加 | 1: そう思う, 2: そう思わない |
| Q35-14 | (博士進学に必要なと考える要件) 民間での雇用条件改善 | 1: そう思う, 2: そう思わない |
| Q37 | インターンシップの経験 | 1: 有, 2: 無 |
| Q46 | 海外留学経験 | 1: 有, 2: 無 |
| Q47-05 | 修士課程の総合満足度 | 1: とても良い~5: 全く良くない |
| Q50 | 1日当たりの平均研究時間 | 1: 1時間未満 2: 1時間以上 3時間未満 3: 3時間以上 5時間未満 4: 5時間以上 8時間未満 5: 8時間以上 12時間未満 6: 12時間以上 15時間未満 7: 15時間以上 |
| Q52 | 査読付き論文投稿数 | 0: 0本, 1: 1本, 2: 2本, 3: 3本以上 |

■構築したデータセットの概要 本研究では特に、「在籍者の分野分類」、「経済的支援に関連する状況や研究活動」、「研究環境に関する状況」、「博士課程進学を検討するために必要と考える条件」と、「実際の進路」との関係性を明らかにすべく、これらに関連する質問への回答を抽出し、データセットとして再構築した。表3には、抽出した回答の本調査における質問番号と各質問内容、そして抽出した回答類型についてまとめた。

BN での分析に際し、回答結果をなるべく簡潔に分類して扱うため、今回のデータセットにおいては数量や自由記述による回答を求めるものについては含めていない。例えば“Q21: 授業料免除”・“Q22: 借入金”については「有」と回答した場合、その金額も回答するようになっているが、今回のデータセットにおいては有無のみを採用し、金額は対象に含めていない。“Q35: 博士進学に必要なと考える要件”については、博士課程進学に必要なと考える要件を複数選択する形式の質問となっていたため、選択肢ごとに選ばれたか、選ばれていないかを「そう思う」「そう思わない」という形で回答されたものとして扱う。また、この Q35 に含まれる選択肢のうち、「その他」は自由記述を伴うため除外した。

また、これらの質問のうち、回答がなされていない個票データについては対象外（リストワイズ）とし、本研究のために再構築したデータセットに含まれるデータの点数は 14,270 点となった。

なお、本研究における分析上の注意点として、Q35 については 01~15 の選択肢のうち 3 つまでしか選択できない回答仕様であることが挙げられる。本研究における DAG の出力でも、Q35 の各ノード間に本仕様に伴う関係が表れる可能性がある。したがって、Q35-01~Q35-14 の間に何らかの関係が見られた場合、これが回答仕様に基づくものか、因果関係の可能性に基づくものか区別することは困難であると考えられ、解釈に注意を要する。

■匿名加工処理について 修士課程調査の個票データは個人情報を含む。そこで個人情報保護法を遵守した形で分析を行うため、事前に匿名加工を行った。具体的には、

- 氏名・年齢・所属大学や住所等分析に不要な情報の削除
- “Q12: 分野分類”について、物理や化学など 45 種類からなるオリジナルの分類を、表 3 の通り学校基本調査に準じた 10 分類に再分類
- “Q52: 投稿論文数”について、3 本以上を一つの層に集約

といった処置を行い、 k -匿名性の観点において $k = 7$ ま

での匿名性を確保している。

4 計算方法・結果

4.1 計算環境・アルゴリズム

ここでは、R version 4.0.4 (2021-02-15) ^{*3} 上で bnlearn パッケージ version 4.7.1 を用いて計算した。本パッケージでは、表 2 に示した例を含む様々なアルゴリズムが利用可能であるが、ここでは代表的なアルゴリズムとして HC・PC・MMHC の 3 種類を採用した。

4.2 計算条件

■**ブラックリストの設定** 計算に際し、そもそも因果の方向として想定できないパスが存在する。例えば、心拍数と歩行速度の関係について、歩行速度から心拍へのパスはあっても、心拍から歩行速度へのパスは考えにくい。今回採用した bnlearn パッケージでは、上記の事例のように、そもそも因果関係がないと考えられるパスは、あらかじめ除外した上で計算できるよう、ブラックリストに記載することができる。

本研究においては以下の通りブラックリストを設定した。

1. “Q12: 分野分類”, “Q16: 社会人経験” は相互の上位下位関係は認めるが、その他の問項目のノードの下に位置することはあり得ない。

2. “Q21: 授業料免除”, “Q22: 借入金”, “Q23: 修了後の進路”, “Q37: インターンシップの経験”, “Q46: 海外留学経験”, “Q47_05: 修士課程の総合満足度”, “Q50: 平均研究時間”, “Q52_01: 論文数” は、上記 Q12, 16 の下位および同レベル内で上位下位に位置することはあるが、その他のノード (Q35_XX 系) の下に来ることはない。

修士課程入学後に定まる論文数や修士課程の満足度が、入学前に選択したはずの専攻分野に影響することは時系列的にあり得ない。同じく、入学後の借入金の有無や、修了後の進路が入学前の社会人経験に影響することもあり得ない。また、例えば論文数やインターンシップ経験が博士進学に必要と考える要件に影響を及ぼす可能性は考えられるが、逆はやや考えにくい。上記のブラックリストはこれらの観点に基づくものであり、妥当と考えられる。

■**検定の条件について** PC アルゴリズム・MMHC アルゴリズムでは、各ノード間の因果関係がありうるか否かについて、条件付き独立性の評価に基づき判断し、枝刈りを行う。本研究では独立性評価は χ^2 検定により行い、有意水準を 1% に設定した。

4.3 計算結果

HC・PC・MMHC の 3 種類で因果探索した結果を、グラフとして図 2 から図 4 に描画した。ここでは比較容易性のためにノードの位置は全グラフで同一とし、エッジ (パス) のみ描き分けている。

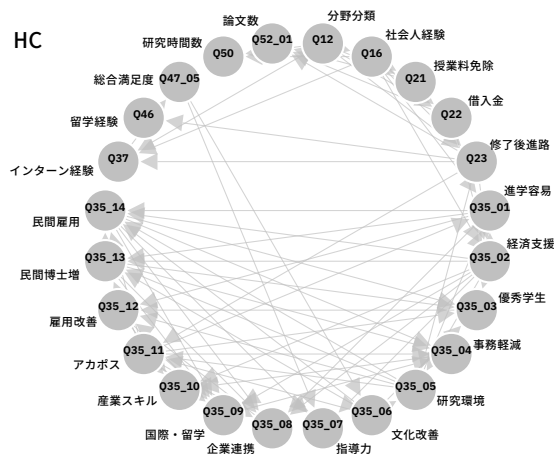


図 2: HC アルゴリズムでの計算結果

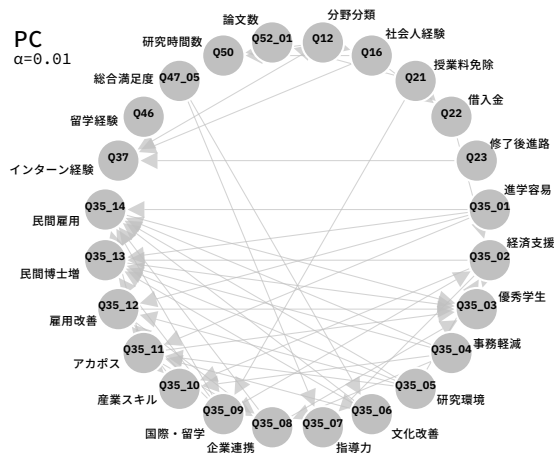


図 3: PC アルゴリズムでの計算結果

結果を見るとアルゴリズムによって具体的パスには差も見られるが、全てに共通するパスも観察できる。

図 2 から図 4 をみると、どのグラフでも “Q12: 分野分類”, “Q16: 社会人経験”, “Q23: 修了後の進路”, から “Q37: インターンシップの経験” へのパス, そして “Q47_05: 修士課程の総合満足度” から “Q35_06: 文化的環境改善”, “Q35_07:PI の指導力向上” へのパスが確認できる。

本稿では 3 つのアルゴリズムで共通するこれらの計算結果に特化して分析・考察を行う。

^{*3} Platform: x86_64-apple-darwin20.3.0 (64-bit)

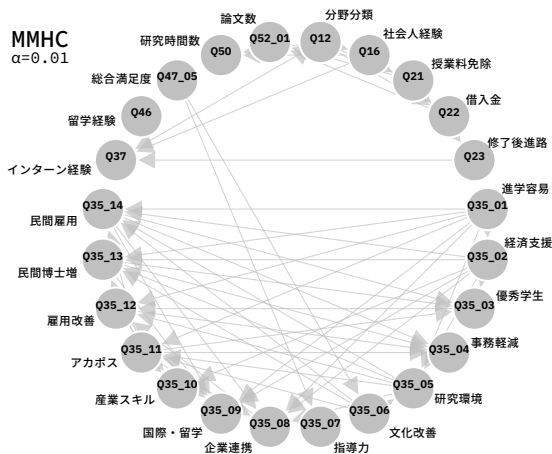


図 4: MMHC アルゴリズムでの計算結果

表 4: Q37 インターン経験 — Q16 社会人経験の CPD

| Q37 \ Q16 | 社会人経験有 | 社会人経験無 |
|-----------|--------|--------|
| インターン経験有 | 0.10 | 0.58 |
| インターン経験無 | 0.90 | 0.42 |

5 考察

5.1 Q37 インターン経験の有無に関する因果推論

“Q37: インターンシップ経験” に関しては、特に“Q12: 分野分類” との関係について、既に修士課程調査の報告書にて定量的な比較についても考察されている [治部 21]。そこで、本稿では“Q16: 社会人経験”，“Q23: 修了後の進路” との関係に絞って、CPD (条件付き確率分布, Conditional Probability Distribution) の値も踏まえながら定性的に議論する。

■“Q16: 社会人経験” との関係 表 4 に示した CPD では、社会人経験がある場合には修士課程でインターンシップを経験しない者が約 9 割で、社会人経験がない場合には半数以上がインターンシップを経験するという結果になっている。インターンシップは社会人を対象外としているプログラムも多く、また時間的な制約等もあるため、特に現役社会人は運用面の理由からこうしたプログラムに参加しづらいといった可能性が考えられる。他にも、社会人経験者は民間企業等の雰囲気をよく理解していることから、修士課程で積極的にインターンシップに行くような動機付けがないという可能性も考えられ、ある程度自然な結果といえる。

■“Q23: 修了後の進路” との関係 表 5 に示した CPD では、「博士課程の進学準備もしくは検討中」の者について、インターンシップを経験しない者が 8 割以上であるのに対し、就職する場合には半数以上がインターンシッ

表 5: Q37 インターン経験 — Q23 修了後の進路の CPD

| Q37 \ Q23 | 博士進学 | 就職 |
|-----------|------|------|
| インターン経験有 | 0.18 | 0.54 |
| インターン経験無 | 0.82 | 0.46 |

表 6: Q35_06 文化改善 — Q47_05 総合満足度の CPD

| Q35_06 \ Q47_05 | 1 | 2 | 3 | 4 | 5 |
|-----------------|------|------|------|------|------|
| そう思う | 0.05 | 0.08 | 0.17 | 0.27 | 0.37 |
| そう思わない | 0.95 | 0.92 | 0.83 | 0.73 | 0.63 |

* Q47_05 の数字は小さいほど満足度が高い。

プを経験するという結果になっている。特に博士課程進学については、はじめから就職をせず博士課程に進学すると決めていれば、インターンシップに行く動機が生じ難いという解釈が考えられる一方、就職については、通常はインターンシップは就職の内定・内々定よりも前に経験するものであり、就職が決定していることによるインターンシップの経験の有無への影響は考えにくい。ゆえに、Q23 と Q37 の間は単純な一方向の因果関係で説明することが困難である。

5.2 Q47_05 修士課程の総合満足度に関する因果推論

■“Q35_06: 文化改善” との関係 表 6 に示した CPD によれば、修士課程における総合的な満足度が低いほど、博士課程進学の要件として、“時に「侮辱的」「抑圧的」「敵対的」な文化的環境が改善されることが重要” だと考える傾向にある。直ちにこの因果関係の可能性を解釈できるわけではないものの、例えば修士課程学生の総合的な満足度の決定要因のひとつに、本人が置かれている研究室の文化的環境があり、この文化的環境に課題を感じている修士課程学生にとって博士課程進学における阻害要因となっている、という仮説が考えられる。

■“Q35_07: PI 指導力” との関係 表 7 に示した CPD によれば、“Q35_06: 文化改善” と同様、修士課程における総合的な満足度が低いほど、博士課程進学の要件として“PI の指導力向上” を重要だと考える傾向にある。こちらも、例えば修士課程学生の総合的な満足度の決定要因のひとつに、指導教員による研究指導の質があり、満足な研究指導を受けられていないと感じている修士課程学生が、博士進学に際して PI の指導力の向上が重要と考えている、という仮説が考えられる。ただし、“Q35_06: 文化改善” の場合と比べて、傾向は緩やかである。

表 7: Q35_07 PI 指導力 — Q47_05 総合満足度の CPD

| Q35_07 \ Q47_05 | 1 | 2 | 3 | 4 | 5 |
|-----------------|------|------|------|------|------|
| そう思う | 0.05 | 0.04 | 0.08 | 0.11 | 0.25 |
| そう思わない | 0.95 | 0.96 | 0.92 | 0.89 | 0.75 |

* Q47_05 の数字は小さいほど満足度が高い。

5.3 本手法での限界・課題

ここまで、試行的な因果探索の結果と解釈について述べた。本節ではデータ面・技術的な限界に伴う問題について述べる。

■調査データに関する課題 まず、本研究の元となるデータはあくまでアンケートである^{*4}。また、悉皆調査ではあるものの、修士課程調査の回答回収率は全体で 13.0% に過ぎず、サンプリングバイアスも想定される。加えて、例えば理学分野では回収率が 30% を超えている一方、保健分野の修士課程では 10% を下回っているなど、分野によっても回収率は大きく異なる。

■アルゴリズムの違いに関する課題 既に示したとおりアルゴリズムによっても出てくる関係性は変わり、またブラックリストの設定内容や有意水準によっても変わることが予想される。本稿で掲載したもの以外にも BN での構造学習のアルゴリズムは存在し、それらの性質も踏まえて、適切な結果をどのように選択していくかは、別途検討が必要である。

■博士課程進学への促進要因・阻害要因 本稿で試行した範囲では、博士課程進学の直接の促進要因や阻害要因の特定までは至らなかった。考えられる要因として、

- Q35_XX 系（博士進学に必要と考える要件）には 3 つまでしか選択できないという拘束条件があり、他の設問との関係性の評価を難しくしている点
- 授業料免除や経済的支援について、これらの単純な有無のみでのカテゴリー分類では、各問との関係性が抽出しづらい点

等が考えられる。データの下処理や因果探索アルゴリズムにかける条件設定において、さらなる工夫がもたらされる。

6 おわりに

本研究では、若手研究者支援の政策領域における新たな領域知識の獲得を目的として、修士課程調査のデータ

を対象に、分析用データセットの構築と BN による因果探索の試行を行った。因果探索のフェーズでは BN の手法による違いも考慮し複数の手法での試行も実施した。

今回の試行からは、修士課程の総合的な満足度と、文化的環境改善や PI の指導力向上の関係性が観察された。文化的環境改善や PI の指導力向上は博士課程進学に必要な要件としても挙げられることもあり、目新しい知見ではないものの、そうした背景の定量的な根拠を得ることができたと言える。

今後、博士課程進学に関する促進・阻害要因の特定や、その他の修士課程在籍者に関する実態を因果の観点から深掘するにあたっては、データ加工やアルゴリズムの条件設定等も期待される。

参考文献

- [Pearl 85] Judea Pearl. : Bayesian Networks: a Model of Self-Activated Memory for Evidential Reasoning. *Proceedings, Cognitive Science Society*, 329–334, 1985. https://ftp.cs.ucla.edu/pub/stat_ser/r43-1985.pdf
- [Shimizu06] S. Shimizu, P. O. Hoyer, A. Hyvärinen, and A. Kerminen : A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7:2003-2030, 2006. <https://www.jmlr.org/papers/volume7/shimizu06a/shimizu06a.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://www.jmlr.org/papers/volume12/shimizu11a/shimizu11a.pdf>
- [治部 21] 治部眞里, 星野利彦: 修士課程 (6 年制学科を含む) 在籍者を起点とした追跡調査 (2020 年度修了 (卒業) 者及び修了 (卒業) 予定者に関する報告). 文部科学省 科学技術・学術政策研究所 調査資料 (*Research Material*), No.310, 2021. <https://doi.org/10.15108/rm310>
- [清水 17] 清水 昌平: 統計的因果探索. 講談社 機械学習プロフェッショナルシリーズ, 2017.
- [高山 21a] 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: EBPM と統計的因果探索・数理モデルの利活用. 研究イノベーション学会 第 36 回年次学術大会 (予稿集), 公演番号 2G02, 2021.
- [高山 21b] 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” を用いた若手研究者支援政策に関する研究. 研究イノベーション学会 第 36 回年次学術大会 (予稿集), 公演番号 2G03, 2021.
- [高山 22a] 高山正行: 科学技術・イノベーション政策研究への統計的因果探索の利活用の新たな可能性 - 博士課程進学率に関する政策論を例として -. 科学技術・学術政策研究所 第 14 回政策研究レビューセミナー, 2022. <https://www.youtube.com/watch?v=GMSWn2abRXw>
- [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: 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>
- [鳥海 18] 鳥海航, 生方裕一, 久野謙也, 岡田幸彦.: 地域健康政策へのベイジアンネットワークの応用. *統計数理*, Vol.66, No.2, pp.267–278, 2018. <https://www.ism.ac.jp/editsec/toukei/pdf/66-2-267.pdf>

^{*4} 自己申告であるため、例えば、実際には社会人経験が「ある」のに「ない」と答える/誤回答することもできる。