

大規模言語モデルを活用した統計的因果探索の新展開

— 統計的因果プロンプティングによる専門知識とデータの融合 —

三内 顕義 | 京都大学理学部特定准教授

国立情報学研究所 大規模言語モデル研究開発センター 客員准教授

高山 正行 | 文部科学省科学技術・学術政策研究所 第1調査研究グループ 研究官

滋賀大学 データサイエンス・AIイノベーション研究推進センター 客員研究員

清水 昌平 | 大阪大学 産業科学研究所 教授

滋賀大学 データサイエンス学系 教授



三内 顕義

京都大学理学部特定准教授。東京大学博士（数理科学）。京都大学数理解析研究所特定助教、理化学研究所革新知能統合研究センター研究員、同上級研究員などを経て2023年より現職。国立情報学研究所大規模言語モデル研究開発センター客員准教授、滋賀大学データサイエンス学部特定准教授、科学技術・学術政策研究所客員研究官などを兼任。

2025年6月に、数学分野でのAIエージェントで世界最高性能（SOTA）を達成。

要旨

本稿は、我々が提案した「統計的因果プロンプティング（SCP）」を解説する。データから因果関係を発見する統計的因果探索（SCD）は多分野で活用されているが、専門知識なしでは正確性に限界があった。この課題に対し、大規模言語モデル（LLM）とSCDを相互参照的に統合する新手法が開発された。SCPは、初期SCD実行、LLMによる知識生成と確率的評価、事前知識を用いたSCD再実行の4段階で構成される。実験では、ベンチマークデータで構造的ハミング距離が縮まり、精度も向上することが確認された。LLMの学習データに含まれない健康診断データでも有効性が確認された。本手法は、統計的証拠と専門知識を融合させる点で画期的だが、計算量に変数の数の2乗に比例する制約がある。また、医療・金融等の重要領域で応用する上では、専門家による検証が不可欠である。今後、LLM技術の発展とともに、より大規模で複雑な因果関係の解明への貢献が期待される。これらの結果はTakayama, et al.(2025)に基づいたものである。

1. はじめに

1.1 因果推論の重要性と現代的課題

私たちは日々、様々な現象の原因と結果を理解しようとしている。なぜ特定の治療法が病気に効くのか、どのような経済政策が景気回復につながるのか、環境汚染がどのように健康に影響するのか。これらの問いに答えるには、単なる相関関係ではなく、真の因果関係を明らかにする必要がある。

近年、ビッグデータの時代を迎え、大量のデータから因果関係を発見する統計的因果探索（Statistical Causal Discovery: SCD）への期待が高まっている。SCDは、観察データから変数間の因果構造を推定する手法であり、ラン

ダム化比較試験が困難な状況でも因果関係の解明を可能にする。

しかし、SCD には重要な課題がある。純粋にデータ駆動的なアプローチでは、統計的には妥当でも現実世界の知識と矛盾する因果関係を導出することがある。例えば、アイスクリームの売上と溺死事故の間に強い相関があっても、両者の間に直接的な因果関係があるわけではない。このような誤った推論を避けるには、分野固有の専門知識、すなわちドメイン知識の活用が不可欠である。

従来、ドメイン知識の組み込みは専門家による手作業に依存していた。しかし、変数が増えるにつれて、すべての変数間の因果関係について専門家が判断することは現実的ではなくなる。さらに、専門家の知識にも偏りや限界があり、最新の科学的知見を常に反映させることは困難である。

1.2 大規模言語モデルの登場と新たな可能性

このような状況に大きな変革をもたらしたのが、大規模言語モデル (Large Language Model: LLM) の登場である。GPT-4 をはじめとする最新の LLM は、膨大な文献から学習した知識を活用し、専門的な質問にも高い精度で回答できるようになった。

LLM の特筆すべき点は、単に情報を記憶しているだけでなく、文脈に応じて推論し、知識を統合できることである。この能力は、因果推論においても大きな可能性を秘めている。例えば、医学的な因果関係について問われれば、解剖学的構造、生理学的メカニズム、疫学的証拠などを総合的に考慮した回答を生成できる。

しかし、LLM を因果推論に活用する試みはまだ始まったばかりである。LLM の出力には不確実性があり、時として事実と異なる情報を生成する「ハルシネーション」と呼ばれる現象も知られている。また、LLM が持つ知識と、特定のデータセットから得られる統計的証拠をどのように統合すべきかも明確ではなかった。

本稿で紹介する我々の研究は、まさにこの課題に正面から取り組んだものである。我々が提案した「統計的因果プロンプティング (Statistical Causal Prompting: SCP)」は、LLM の持つドメイン知識と SCD による統計的証拠を相互参照的に統合する画期的な手法である。この手法により、専門知識を考慮しつつ、データに基づいた信頼性の高い因果推論が可能になる。

以下では、この革新的なアプローチの詳細を解説し、その可能性と課題、そして今後の展望について論じていく。

2. 原論文 Takayama, et al.(2025)の概要

2.1 研究の背景と目的

我々の研究は、統計的因果探索が抱える根本的な課題に着目している。従来の SCD 手法は、データの統計的傾向のみに基づいて因果関係を推定するため、しばしば現実と乖離した結果を生む。例えば、病院の数と死亡率の間に正の相

関があったとしても、「病院が増えると死亡率が上がる」という因果関係は明らかに誤りである。このような問題を解決するには、「病気の人が多い地域に病院が建設される」という背景知識が必要となる。

研究の目的は、LLM が持つ広範な知識を活用して、このようなドメイン知識を体系的に SCD に組み込む方法を開発することである。特に重要なのは、LLM の知識とデータから得られる統計的証拠を単に組み合わせるのではなく、両者が相互に情報を提供し合う枠組みを構築した点である。

2.2 提案手法の概要

SCP の核心は、4 段階の反復的プロセスにある。

第 1 段階では、通常の SCD アルゴリズムを用いて、事前知識なしに因果グラフを推定する。この初期結果は、データに内在する統計的傾向を反映している。

第 2 段階では、LLM に対して変数間の因果関係について問いかける。ここで重要なのは、単に「A は B の原因か」と尋ねるのではなく、第 1 段階で得られた統計的情報も提示することである。例えば、「統計解析では A から B への因果関係が示唆されていますが、専門知識の観点からこれは妥当でしょうか」というかたちで問いかける。

第 3 段階では、LLM が生成した詳細な説明を基に、各因果関係の存在確率を定量的に評価する。因果関係の存在を LLM に「Yes」か「No」で回答させるかたちで問い、その出力確率を測定することで、因果関係の確信度を数値化する。

第 4 段階では、LLM から得られた確率的評価を事前知識行列に変換し、これを制約条件として用いて、再度 SCD を実行する。この際、統計的証拠と専門知識の両方を考慮した、より信頼性の高い因果グラフが得られる。

3. 主要な理論・手法の解説

3.1 統計的因果探索の基礎

統計的因果プロンプティング (SCP) を理解するには、まず基盤となる統計的因果探索 (SCD) の主要手法を理解する必要がある。我々は、代表的な三つのアルゴリズムを採用している。

3.1.1 PC アルゴリズム

PC アルゴリズムは、ピーター・スパーティスとクラーク・グリマーによって開発された、制約ベースの因果探索手法である。その名前は開発者の頭文字に由来する。

このアルゴリズムの基本原理は、条件付き独立性の検定である。二つの変数が与えられたとき、他の変数を条件として独立かどうかを統計的に検定する。例えば、気温とアイスクリームの売上は相関するが、季節を条件とすると独立になる可能性がある。この場合、気温と売上の間に直接的な因果関係はなく、

両者は季節という共通原因を持つと推定できる。

PC アルゴリズムは、まず完全グラフから始めて、条件付き独立性が成立するペアのエッジを削除していく。その後、向きのない辺で構成されるグラフ構造（スケルトン）が得られたら、有向非巡回グラフ（DAG）の仮定等に基づく向きづけルール（オリエンテーションルール）を適用し、可能な範囲で因果の向きを特定していく。このプロセスにより、最終的に、データと統計的に整合する因果グラフが得られる。ただし、観察データのみからは、オリエンテーションルールで因果の方向を完全に特定できない場合もあり、その場合は無向辺を残した、部分的有向非巡回グラフ（CPDAG）として表現される。

3.1.2 Exact Search アルゴリズム

Exact Search は、スコアベースの手法であり、可能なすべての DAGの中から、データに最も適合するものを探索する。我々は、A* アルゴリズムを用いた、ヒューリスティックなアプローチを導入している比較的効率的な実装を採用している。

この手法では、候補となる各因果グラフに対してベイズ情報量基準（BIC）などのスコアを計算し、そのスコアが最適となる因果グラフを選択する。BIC は、モデルの適合度とモデルの複雑さのバランスを考慮した指標であり、過学習を防ぐ効果がある。理論的には、サンプルサイズが十分大きければ、真の因果グラフを一致推定できることが知られている。

スコアベースの手法の中でも、Exact Search の利点は、グローバルに最適な解を保証できることである。一方、変数の数が増えると計算量が指数的に増加するため、大規模なネットワークには適用が困難という課題もある。

3.1.3 DirectLiNGAM アルゴリズム

LiNGAM（Linear Non-Gaussian Acyclic Model）は、清水昌平氏らによって開発された、日本発の革新的な因果探索手法である（Shimizu et al.(2006)）。DirectLiNGAM はその効率的な実装版である。

従来の統計的手法の多くは、モデルの誤差項が正規分布に従うことを仮定していたが、LiNGAM は非ガウス性を仮定し、非ガウス分布の場合には、観察データのみから因果の方向を一意に特定できるという画期的な性質を持つ。

DirectLiNGAM は、変数の因果的順序を逐次的に同定していく。各ステップで、他の変数から最も独立な変数を選び、それを最上流の変数として確定する。この過程を繰り返し、回帰分析を行うことで、完全な因果グラフを構築し、因果係数の形で因果効果を推定することができる。

3.2 統計的因果プロンプティング（SCP）の詳細

SCP の革新性は、LLM へのプロンプト設計にある。単に LLM に因果関係を尋ねるのではなく、統計的証拠を含めた文脈を提供することで、より信頼性の高い推論を引き出す。

3.2.1 Generated Knowledge Promptingの活用

Generated Knowledge Prompting は、LLM に直接答えを求めるのではなく、まず関連する知識を生成させてから、生成された知識も活用して判断を行わせる手法である。我々は、この手法を因果推論に適用している。

具体的には、変数 A と B の因果関係について、まず「A が B に影響を与えるメカニズムについて、専門知識に基づいて説明してください」と問いかける。LLM は、関連する理論、既知のメカニズム、実証研究の知見などを詳細に説明する。この過程で、LLM は自身の知識をより着実に活用できるようになり、より深い推論が可能になる。

3.2.2 Zero-Shot Chain-of-Thoughtによる推論誘導

Chain-of-Thought (CoT) は、LLM に段階的な推論過程を明示させることで、複雑な問題の解決能力を向上させる手法である。Zero-Shot CoT は、特別な例示なしにこれを実現する。

SCP では、「統計解析の結果も踏まえて、この因果関係の妥当性を、専門家の見地から評価してください」という形で、LLM に推論過程の明示を促す。LLM は、統計的証拠、理論的整合性、既知の事実との一致性などを順に検討し、最終的な判断に至る。

3.2.3 プロンプトパターンの設計思想

我々は、五つのプロンプトパターンを設計し、それぞれの効果を検証している。

パターン 0 (ベースライン) では、統計的情報を含めず、純粋に LLM が有するドメイン知識のみで判断させる。これは、LLM の基本的な推論能力を測る基準となる。

パターン 1 では、初期 SCD で得られた因果グラフの構造を提示する。「統計解析では、A から B への因果関係が示唆されています」という形で、因果の方向性の情報を含める。

パターン 2 では、ブートストラップ法で計算した因果関係の出現確率を追加する。「1000 回のリサンプリングで、この因果関係が 75% の確率で検出されました」といった定量的情報を提供する。

また、構造方程式を仮定して因果係数も出力する DirectLiNGAM の場合のみではあるが、パターン 3 では推定された因果係数の値、パターン 4 ではパターン 2・3 の両方の値を、というように、さらに詳細な統計量も含める。これにより、因果関係の強さも LLM に提供される。

3.2.4 トークン生成確率の測定と信頼度評価

LLM の出力の信頼度を定量化するため、トークン生成確率を活用する。変数間に因果性があるか否かという質問に対し LLM が「Yes」または「No」と回答する際の対数確率を取得し、これを因果関係の存在確率として解釈する。

ただし、LLM の出力確率にも揺らぎがあることが知られており、同じプロンプトでも、実行のたびに異なる確率が出力される。この問題に対処するため、我々は同じ質問を 5 回繰り返し、平均確率を採用している。この手法に

より、より再現性の高い評価が可能になる。

3.3 LLMとSCDの統合メカニズム

3.3.1 事前知識行列の生成プロセス

LLM から得られた確率的評価を、SCD アルゴリズムが利用できる形式に変換する必要がある。我々は、確率に閾値を設定し、それを超える因果関係を事前知識として採用する方法を提案している。

具体的には、 n 個の変数に対しての $n \times n$ 行列を作成する。行列の (i, j) 要素は、変数から変数への因果関係の有無を表す。LLM の出力確率が閾値（例えば 0.05）を下回る場合は、その辺は絶対に現れないものとして 0 とする。

3.3.2 アルゴリズムの計算複雑度とスケーラビリティ

SCP の計算量は、主に LLM への問い合わせ回数によって決まる。 n 個の変数に対して、 $n(n-1)$ 個の因果関係を評価する必要があるため、計算量は $O(n^2)$ となる。さらに、各関係について 5 回の測定を行うため、実際の計算時間は $5n(n-1)$ 回の LLM 呼び出し回数に比例する。

この二次の計算複雑度は、大規模ネットワークへの適用において課題となる。例えば、100 変数のネットワークでは約 50,000 回の LLM 呼び出しが必要となり、API 制限やコストを考慮すると現実的ではない。したがって、本手法が適用可能な問題は、変数の数（モデルの規模、考察する範囲）に制約される。

3.4 手法の技術的工夫と革新性

3.4.1 相互参照的な性能向上の仕組み

SCP の最も革新的な点は、LLM と SCD が相互に情報を提供し合うことで、両者の性能の向上が期待される点である。SCD の結果を LLM に提示することで、LLM はデータの統計的傾向を考慮した推論ができる。一方、LLM の知識を SCD に組み込むことで、現実と整合的な因果グラフが得られる。

この相互参照は、人間の専門家が統計解析結果を解釈する過程に似ている。専門家は、統計的に有意な関係を見たとき、それが既知の理論やメカニズムと整合するかを検討する。同時に、統計的証拠が強ければ、既存の理論を修正することもある。SCP は、このような高度な推論プロセスを自動化している。

3.4.2 統計的証拠とドメイン知識の融合方法

従来の研究では、統計的手法はドメイン知識に基づくモデル構築の議論とは独立に発展してきた。SCP は、両者を LLM の出力の確率を介することで統一的に扱うことに成功している。

この融合により、データが少ない場合はドメイン知識に基づいて補完を試みる一方、データが豊富な場合はその統計的根拠に基づいてドメイン知識の確からしさについての定量的な議論にも踏み込むという適応的な推論も可能にな

る。これは、ベイズ統計における事前分布と尤度の関係に類似しており、理論的にも妥当なアプローチである。

4. 実証結果とその含意

4.1 ベンチマークデータでの検証

我々は、手法の有効性を検証するため、因果探索の研究で広く使用されるベンチマークデータセットを用いた実験を行っている。これらのデータセットは、因果関係についてある程度、専門家の共通見解が得られているか、あるいは一般常識になっているような変数の組み合わせで、現実世界で実際に取得されたデータであり、因果探索アルゴリズムの性能評価にもよく使われるものである。ベンチマークになる因果関係 (ground truth) があるため、因果探索した結果がそのベンチマークにどれだけ近づいているかをもとに、手法の性能を客観的に評価できる。本稿ではそのうち、Auto MPG データ、DWD 気候データでの検証事例を紹介する。

4.1.1 Auto MPGデータ：小規模ネットワークでの安定性

Auto MPG データセットからは、自動車の燃費に関する、重量 (weight)、馬力 (horsepower)、加速性能 (acceleration)、総排気量 (displacement)、燃費 (miles per gallon) の計五つの連続変数を抽出した。このデータセットは変数の数が少ないため、SCP の基本的な動作を検証するのに適している。

実験結果では、すべてのプロンプトパターンで安定した性能向上が観察された。構造的ハミング距離 (SHD) は2つの有向グラフ間の構造的な違いを測る指標である。真の因果性を示す有向グラフと推計結果を有向グラフ化したものを比較することで、推計精度を確認することができる。同指標では、事前知識なしのベースラインと比較して50%以上改善した。興味深いことに、異なるプロンプトパターン間での性能差は小さく、これは変数の数が少ない場合のSCPの頑健性を示している。また、事前知識なしのSCDでは、燃費→重量、燃費→馬力といった、不自然な辺を出力する傾向があったが、SCPにより、「重い車は燃費が悪い」「馬力大きい車は燃費が悪い」といった、物理的にも自然な因果関係に、正しくSCDの結果を誘導できることが確認された。

4.1.2 DWD気候データ：中規模での効果検証

DWD データセットは、ドイツ気象局の気候データから作成された6変数のデータセットであり、緯度 (latitude)、経度 (longitude)、高度 (altitude)、気温 (temperature)、日照時間 (sunshine)、降水量 (precipitation) などの気象に関連する変数から構成される。

このデータセットでは、SCPの効果がより顕著に現れた。特にDirectLiNGAMを用いた場合、特にパターン2（ブートストラップ確率を含む）で最大60%のSHD改善が見られた。これは、統計的情報がLLMの推論を大きく向上させることを示している。

事前知識なしのSCDでは、気温→高度など、地理的に定まっている変数

が他の変数から影響を受けるような不自然な辺が出力されることがある一方、SCP を経ることで、事前知識ありの SCD では緯度・経度・高度が他の変数から影響を受けるといった有向グラフの辺が現れることはなくなり、むしろ「高度が高いと気温が下がる」といった、気象学的にも自然な辺が、SCD に適切に出力されるようになった。

4.2 未公開健康診断データでの実証

ベンチマークデータでの成功は重要だが、公開された著名なベンチマークであり実証研究で数多く利用されているため、LLM が単に学習データを記憶している可能性を完全には排除できない。そこで我々は、GPT-4 の学習データに含まれていない日本の健康診断データを用いた実験を行った。

4.2.1 データの特性と実験設計

使用されたデータは、年齢、HbA1c（糖化ヘモグロビン）、肥満度、中性脂肪値、腹囲、収縮期血圧、拡張期血圧の7変数で構成している。これらは糖尿病リスクに関連する重要な指標である。データは特定の者しかアクセスできないように厳重に管理されており、当然、インターネットにも開示されていないため、LLM が事前に学習している可能性は極めて低い。

実験では、全データから無作為に抽出したサブセットのうち、統計的バイアスを含む結果を、意図的に検証に採用した。これにより、純粋な統計的手法のみでは正しい因果関係を発見しにくい、より現実的な条件での検証が可能となった。

4.2.2 LLMによる因果探索結果の改善の確認

まず、事前知識を組み込まない初期の SCD では、「HbA1c が年齢に影響する」といった、非常に不自然な結果が出力されていた。一方、SCP を進めると、「年齢が上がると HbA1c が上昇する」といった、医学的に確立された因果関係を正しく認識し、その推論結果を事前知識として DirectLiNGAM で再度因果探索した結果、年齢→HbA1c 経路が正しく推定されるなど、医学的により妥当な因果グラフが得られた。

このように、LLM は未知のデータセットに対しても適切な推論を示した。また、この結果は、SCP が実際の医療データの因果分析においても、有用である可能性を示唆している。特に、観察研究では避けられない選択バイアスや交絡の問題によって、データセットに何らかの偏りが発生してしまう場合でも、LLM の知識が補正効果を持つことが確認された。

4.3 結果の統計的・実務的解釈

4.3.1 性能評価指標の改善とその意味

特にベンチマークデータを活用した実験全体を通じて、複数の評価指標で一貫した改善が見られた。例えば構造的ハミング距離（SHD）の改善は、推定された因果グラフが真の構造により近づいたことを意味する。精度（Precision）

の向上は、誤った因果関係の削減を、精度と再現率（Recall）の両者を総合評価する F1 スコアの改善は、全体的なバランスの向上を示している。

また、統計的妥当性の指標である CFI（比較適合度指標）や RMSEA（近似の二乗平方平均誤差）の極端な悪化はほとんど観測されず、これは SCP により得られた因果モデルがデータにより良く適合することを示している。さらに、BIC の減少は、モデルの複雑さを考慮してもなお、改善されたモデルが優れていることを意味する。

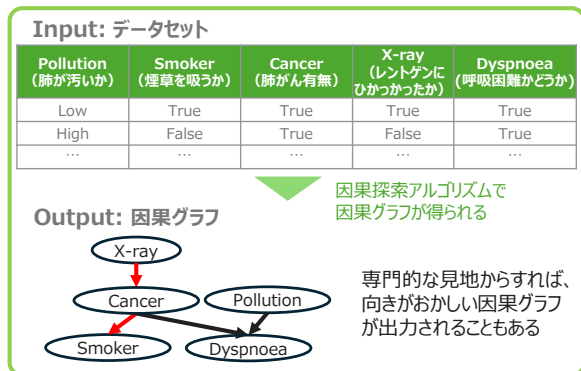
4.3.2 プロンプトパターンの効果分析

プロンプトパターンの分析からは、「適度な」統計情報の提供が重要であることが明らかになった。例えば、パターン 2（ブートストラップ確率のみ）は多くの場合で良好な結果を示したが、パターン 4（すべての統計情報を含む）では必ずしも最良ではなかった。

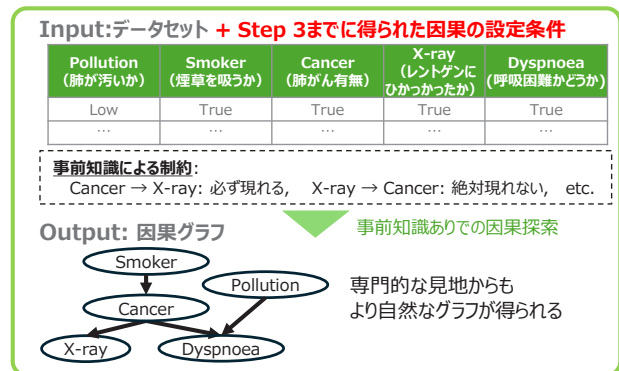
これは、情報過多による LLM の判断の揺らぎ、あるいは統計的詳細がドメイン知識の活用を妨げることもあるという可能性を示唆している。実務的には、使用する SCD アルゴリズムと問題の特性に応じて、適切なプロンプトパターンを選択することが重要である。

図表 1 統計的因果プロンプティング(SCP)の概念図

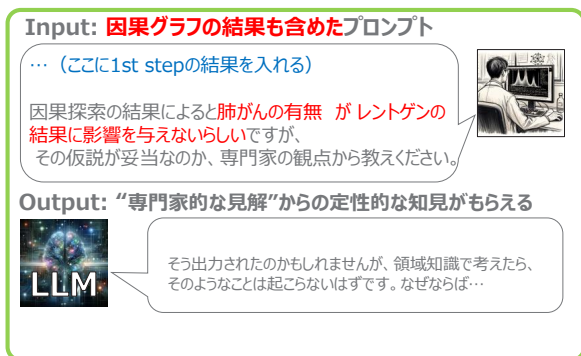
1st step: データ駆動の因果探索（事前知識なし）



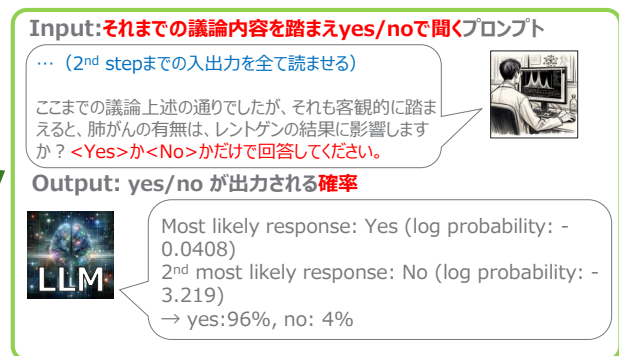
4th step: 事前知識ありで再度因果探索を実行



2nd step: 因果探索の結果の妥当性をLLMに問う



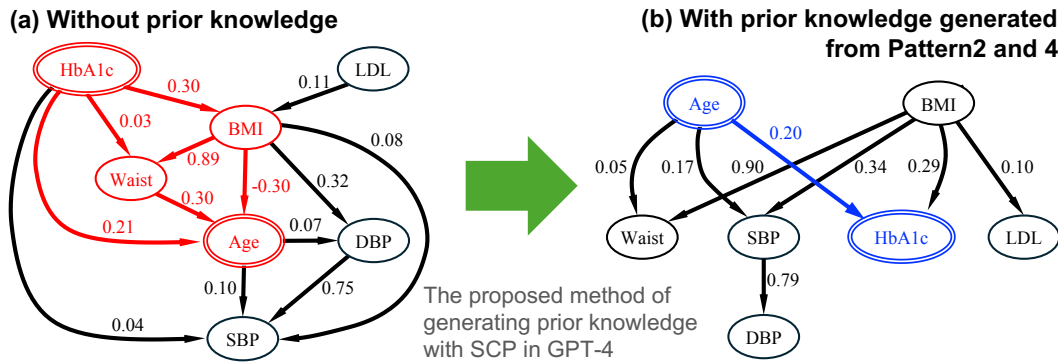
3rd step: LLMがそこまでの議論から客観的/確率的に判断



出所) 筆者作成

また、図表2からは、LLM が有する情報の活用により、ヘルスケアデータからの因果探索で現れた年齢（age）への矢印（論理的に存在しない）が適切に排除されていることがわかる。

図表2 ヘルスケアデータにおける因果グラフの比較



出所) Takayama et al. (2025)

5. 政策的・実務的示唆

SCP の開発は、様々な分野での因果推論の実践に大きな影響を与える可能性がある。ここでは、主要な応用分野における活用方法と、実装時の課題・対策について論じる。

5.1 医療・健康科学への応用

5.1.1 疾患要因の探索への活用

医療分野では、疾患の発症メカニズムや危険因子の特定が重要な課題である。SCP は、電子カルテや健康診断データから、より信頼性の高い因果関係を発見するのに役立つ可能性がある。

例えば、生活習慣病の発症要因を分析する際、単純な統計解析では「運動不足→肥満→糖尿病」という直線的な関係しか捉えられないことが多い。しかし、従来の因果探索に加え、SCP を用いることで、現実的には同時にデータの取得の難しい遺伝的要因、環境要因、行動要因の複雑な相互作用も領域知識の観点から考慮した、より現実的な因果モデルを構築できる可能性もある。

特に近年の LLM は、追加学習を行わせなくとも、事前学習の段階で、その学習データのナレッジ・カットオフとなる時点までの医学文献の知識も学習しているため、新しく発見された疾患メカニズムや、稀な疾患についても推論に反映が可能になると期待される。これにより、データ数が限られる希少疾患の研究においても、有用な知見が得られる可能性がある。

5.1.2 予防医学での因果関係理解

予防医学の観点からは、介入可能な要因の特定が重要である。従来の SCD に加え、SCP を活用することで、どの要因に介入すれば最も効果的に疾患を

予防できるかを、統計的観点と領域知識の両面から明らかにするのに役立つ。例えば、早期介入により薬剤使用を回避できる可能性のある群を特定することで、医療費の削減と患者の QOL 向上の両立が期待できる。

5.1.3 実装時の注意点

医療分野で SCP を実装する際は、SCD を応用した場合の結果解釈と同様、特に慎重な検証が必要である。LLM の推論結果は、ハルシネーションに由来する致命的な誤りがないことを確認するために、必ず医学専門家によるレビューを経るべきである。また、個人情報保護の観点から、取り扱いに配慮が必要なデータは、匿名化や、LLM 利用時に、誤って機微な内容がプロンプトに打ち込まれないようにするための予防的工夫も必要である。

さらに、医療分野特有の倫理的配慮も重要である。因果関係の誤認識が患者の治療方針に影響を与える可能性があるため、SCP の結果は意思決定支援ツールとして位置づけ、最終的な判断は医療従事者が行うべきである。

5.2 経済・金融分野での活用

5.2.1 市場メカニズムの解明

金融市場では、複雑な要因が相互に影響し合っている。従来の SCD に加え、SCP を用いることで、市場変動の背後にある因果メカニズムをより深く理解できる可能性がある。

例えば、為替レート、金利、株価指数の関係を分析する際、従来の計量経済学的手法では見逃されがちな間接的な因果経路を、SCD により発見できる可能性がある。加えて経済理論や過去の市場動向に関する知識を持っている LLM の活用により、統計的に検出された関係が経済学的に妥当かどうかを判断することも可能であり、真に新しい因果経路についても領域知識から議論することができる。

5.2.2 リスク要因の特定

リスク管理において、真の因果関係の理解は極めて重要である。相関関係に基づくリスク評価では、見かけ上の関係に惑わされる危険がある。SCD に加え LLM を活用する SCP は、真のリスク要因を領域知識の観点からも特定し、より効果的なリスク管理戦略の立案を支援する。

例えば、信用リスク評価では、顧客属性と債務不履行の関係を分析する際、社会経済的な背景知識を考慮することで、より公平で正確なモデルを構築できる。これは、人種や住所により与信の可否が左右されるといったアルゴリズムックバイアスの問題を軽減する上でも重要である。

5.2.3 政策効果の事前評価

経済政策の立案において、政策変更が経済に与える影響を事前に評価することは重要である。SCD に加え、LLM を活用した SCP を用いることで、過去のデータから政策変数と経済指標の因果関係を領域知識からも自然な形で推定

し、政策効果をより正確に予測できる。

5.3 実装における課題と対策

5.3.1 計算資源の制約

SCPの実装における最大の課題の一つは、計算コストである。変数の数の二乗に比例してLLM呼び出し回数が増加するため、大規模な問題では実行時間とコストが問題となる。

この課題に対しては、例えば段階的なアプローチが有効である。まず、ドメイン知識や予備的な統計解析に基づいて変数を絞り込み、重要な変数間の関係に焦点を当てる。また、バッチ処理や並列化により、効率的な実行を図ることも重要である。

5.3.2 専門家による検証の必要性

LLMの出力は完全ではなく、時として誤った推論を行う可能性がある。特に、専門性の高い分野や最新の知見が関わる場合は、注意が必要である。

実務では、LLMの推論結果を専門家がレビューする体制を構築すべきである。SCPの結果を可視化し、推論過程を透明化することで、専門家が効率的に検証できる環境を整える必要がある。

5.3.3 継続的な改善プロセス

LLM技術は急速に進化しており、新しいモデルが次々と登場している。SCPの性能を最大化するには、最新のLLMを活用し、プロンプト設計を継続的に改善する必要がある。

また、特定分野に特化したファインチューニングや、RAG (Retrieval-Augmented Generation) 技術の活用により、より専門的で正確な推論が可能になる。組織内での知識蓄積と共有の仕組みも重要である。

5.4 責任あるAI活用に向けて

5.4.1 透明性の確保

SCPを意思決定に活用する際は、推論過程の透明性が重要である。なぜその因果関係が推定されたのか、LLMがどのような知識に基づいて判断したのかを明確にする必要がある。

プロンプトと応答の記録を保存し、監査可能な形で管理することが推奨される。これにより、結果の再現性を確保し、問題が生じた場合の原因究明も可能になる。

5.4.2 倫理的配慮

因果推論の結果は、政策決定や個人の処遇に影響を与える可能性がある。特に、社会的に敏感な属性（性別、人種、社会経済的地位など）が関わる場合は、慎重な扱いが必要である。

SCP を用いる際も、公平性と説明責任を重視し、結果の解釈と適用において倫理的ガイドラインに従うべきである。また、因果関係の推定が差別や偏見を助長しないよう、継続的なモニタリングが必要である。

6. 今後の展望と課題

SCP は因果推論の新たな地平を開いたが、まだ発展の初期段階にある。ここでは、技術的・理論的な改善の方向性と、社会実装に向けた展望を議論する。

6.1 技術的限界と改善の方向性

6.1.1 大規模ネットワークへの対応

現在の SCP の最大の制約は、計算量が $O(n^2)$ であることである。例えば 100 変数を超えるような大規模ネットワークでは、現実的な時間とコストでの実行が困難である。

この問題に対する有望なアプローチの一つは、階層的な因果探索である。まず変数をグループ化し、グループ間の因果関係を推定した後、各グループ内の詳細な関係を分析する。このようなトップダウンアプローチにより、計算量を大幅に削減できる可能性がある。

また、因果関係の疎性を仮定し、重要な変数ペアのみを選択的に評価する手法も検討に値する。ドメイン知識や予備的な統計解析に基づいて、評価すべき変数ペアを事前にスクリーニングすることで、LLM 呼び出し回数を削減できる。

6.1.2 より高度なプロンプト設計

現在のプロンプト設計は、主に経験的な試行錯誤に基づいている。今後は、日々発展を続けているプロンプト技術を様々に取り入れ、因果推論の目的に応じて最適なプロンプト設計を追究する必要がある。

特に重要なのは、LLM の推論過程をより細かく制御する方法の開発である。例えば、因果関係の時間的順序、介入可能性、交絡要因の存在など、因果推論の重要な概念を明示的にプロンプトに組み込むことで、より精密な推論が可能になるだろう。

また、Few-shot 学習や Chain-of-Thought の発展形など、新しいプロンプティング技術の活用も期待される。分野特有の因果推論パターンを学習させることで、専門性の高い推論が可能になる。

6.1.3 複数LLMのアンサンブル

単一の LLM に依存するのではなく、複数の LLM の推論結果を統合するアンサンブル手法の開発も重要である。異なる LLM は異なる強みを持つため、それらを組み合わせることで、より頑健で信頼性の高い推論が可能になる。

例えば、一般的な知識に強い LLM、医学に特化した LLM、統計的推論に優れた LLM を組み合わせることで、それぞれの専門性を活かした総合的な判断が可能になる。投票メカニズムやベイズ的統合など、複数の推論結果を統合する理論的枠組みの構築が必要である。

6.2 理論的発展の可能性

6.2.1 因果推論理論への貢献

SCP は、統計的手法と知識ベース手法を融合する新しいパラダイムを提示している。この枠組みは、因果推論の理論的発展にも寄与する可能性がある。

特に興味深いのは、LLM の確率的出力を因果推論の不確実性と結びつける理論の構築である。LLM の確信度を事前分布として解釈し、データから得られる尤度と組み合わせることで、ベイズ的因果推論の新しい形式を確立できる可能性がある。

6.2.2 新しい評価指標の開発

現在の評価指標は、主に構造的な一致度に焦点を当てている。しかし、因果推論の実用性を評価するには、より多面的な指標が必要である。

例えば、推定された因果関係の介入可能性、政策的重要性、予測精度への寄与度など、実用的観点からの評価指標の開発が求められる。また、LLM の推論の質を評価する指標も必要である。推論の論理的一貫性、既知の事実との整合性、不確実性の適切な表現など、多角的な評価が重要である。

6.2.3 他の機械学習手法との統合

SCP は、他の機械学習手法と組み合わせることで、さらなる発展が期待できる。例えば、深層学習による特徴抽出と SCP を組み合わせることで、高次元データからの因果発見が可能になるかもしれない。

また、強化学習の枠組みで SCP を活用することで、介入の効果を学習しながら因果構造を発見する、より動的なアプローチも考えられる。これは、実験計画法と因果推論を統合する新しい研究方向を示唆している。

6.3 社会実装に向けたロードマップ

6.3.1 段階的な導入戦略

SCP の社会実装は、段階的に進めることが重要である。まず、リスクの低い領域での小規模な実証実験から始め、徐々に適用範囲を拡大していくべきである。

例えば、第 1 段階では、研究機関や大学での基礎研究への活用から始める。ここでは、手法の改善と検証に重点を置く。第 2 段階では、企業の研究開発部門での活用を進める。製品開発や品質管理など、限定的な領域での応用を試みる。第 3 段階では、より広範な実務への展開を図る。その際、医療、金融、政策立案など、社会的影響の大きい分野への適用を慎重に進める。こうした段

階的な拡張においては、各段階について、フィードバックに基づく改良を繰り返し、次の段階での実証に進むかどうかを判断していく。

6.3.2 産学連携の推進

SCP の発展には、学术界と産業界の密接な連携が不可欠である。大学や研究機関が理論的基盤を提供し、企業が実データと実用的課題を提供することで、相乗効果が期待できる。

共同研究プロジェクトの設立、データ共有の枠組み構築、人材交流の促進など、具体的な連携メカニズムの整備が必要である。また、成功事例の共有と普及も重要であり、ワークショップやカンファレンスの開催が推奨される。

6.3.3 国際標準化への取り組み

SCP が広く採用されるためには、手法の標準化が重要である。評価基準、実装ガイドライン、倫理的指針など、国際的に合意された標準の策定が必要である。

また、異なる言語や文化圏での SCP の適用可能性を検証することも重要である。LLM の多言語対応能力を活用しつつ、各地域の文脈に適したカスタマイズを行う必要がある。国際的な研究コンソーシアムの形成により、グローバルな知見の集約と共有を進めることが期待される。

7. おわりに

本稿では、我々が提案した SCP について、その理論的基礎から実用的応用まで幅広く解説してきた。SCP は、大規模言語モデルの知識と統計的因果探索の手法を革新的な方法で統合し、因果推論の新たな可能性を切り開いた。

7.1 本研究の意義の総括

SCP の最も重要な貢献は、これまで独立に発展してきた二つのアプローチ、すなわちデータ駆動的な統計手法と知識ベースの推論を、実用的な形で融合したことである。この統合により、それぞれの手法を相互に補完し、より信頼性の高い因果推論が可能になった。

実証研究では、既知のベンチマークデータセットにおいて顕著な性能向上が確認された。特に重要なのは、LLM の学習データに含まれていない健康診断データでも有効性が実証されたことである。これは、SCP が単なる記憶の再現ではなく、真の知識統合による推論を実現していることを示している。

また、本研究は手法の限界についても率直に議論し、計算量の問題、LLM の不確実性、実装上の課題など、今後解決すべき点を明確にした点も付記しておく。

7.2 因果推論研究の新時代への展望

SCP の登場は、因果推論研究の新しい時代の幕開けを告げている。AI と

人間の知識が協調して複雑な問題に取り組む時代が到来したのである。今後、LLM 技術のさらなる発展により、より高度で専門的な推論が可能になることが期待される。

しかし、技術の進歩と同時に、その責任ある使用についても真剣に考える必要がある。因果推論の結果は、医療、経済、社会政策など、人々の生活に直接影響を与える分野で使用される。したがって、技術の開発と並行して、倫理的ガイドラインの整備、専門家による検証体制の構築、透明性の確保など、社会的な枠組みの整備も不可欠である。

また、SCP の普及には、研究者、実務家、政策立案者など、様々なステークホルダーの協力が必要である。学際的な対話を通じて、技術の可能性と限界について共通理解を形成し、適切な活用方法を模索していくことが重要である。

7.3 読者へのメッセージ

因果関係の理解は、科学の根幹をなす営みである。なぜ物事が起こるのか、どのように介入すれば望ましい結果が得られるのか。これらの問いに答えることは、人類の知的探求の中心的課題であり続けてきた。

SCP は、この古くて新しい問題に対する革新的なアプローチを提供している。しかし、これは決して万能の解決策ではない。むしろ、人間の専門知識と AI の計算能力を組み合わせることで、より深い理解に到達しようとする試みである。

読者の皆様には、この新しい技術の可能性に期待しつつも、批判的な視点を持って接していただきたい。特に、自身の専門分野で SCP をどのように活用できるか、どのような課題があるか、積極的に検討していただければ幸いである。技術の真の価値は、それを使う人々の創造性と責任感によって決まるのである。

最後に、統計的因果探索と大規模言語モデルに関する日本の研究者に敬意を表するとともに、この分野のさらなる発展を期待したい。因果推論の新時代において、日本の研究者が重要な貢献を果たしていることは、誠に喜ばしいことである。今後も、国際的な協力のもと、この革新的な研究が人類の知識の前進に寄与することを願ってやまない。

参考文献

- Shimizu, S., Hoyer, P. O., Hyvärinen, A., & Kerminen, A. (2006). A Linear Non-Gaussian Acyclic Model for Causal Discovery. *Journal of Machine Learning Research*, 7, 2003-2030.
- Takayama, M., Okuda, T., Pham, T., Ikenoue, T., Fukuma, S., Shimizu, S., & Sannai, A. (2025). Integrating Large Language Models in Causal Discovery: A Statistical Causal Approach. *Transactions on Machine Learning Research*, 10 May 2025.