

DAG-NoCurl に基づく効率的な潜在因子間の因果探索手法の開発

橋本 捷矢

変数間の因果関係を扱う因果推論分野においては、多くの場合、因果関係を有向辺で表した因果グラフを用いて分析が行われる。因果グラフにおける因果構造は、該当分野の専門家による事前知識などを用いて記述されることが多いが、因果グラフが未知である場合、得られたデータから真の因果グラフを復元する必要がある。このような状況において用いられるのが因果探索手法である。因果探索とは、得られた観測データからそのデータを生成するもととなった因果グラフを推定する分析手法である。多くの因果探索手法は、観測変数間の因果グラフを推定することを目的としている。しかし、心理学や社会学など、いくつかの分野では、観測変数間ではなく、潜在因子間の因果構造に興味がある場合がある。潜在因子間の因果探索は、そのような観測データから直接得られることができない潜在因子間の因果構造を推定する手法である。

潜在因子間の因果探索手法の 1 つに LiNA(Linear Non-Gaussian acyclic model for Latent Factors)がある。LiNA は誤差項が非ガウス分布に従う線形の構造方程式モデルを仮定することで、潜在因子間の因果構造を表す有向非巡回グラフ(Directed Acyclic Graph, DAG)を一意的に推定することができる手法である。LiNA のアルゴリズムは、観測変数をもとに潜在因子を見つける測定モデルの推定フェーズと、見つけられた潜在因子間の因果構造を推定する構造モデルの推定フェーズの 2 つにわけられる。LiNA は構造モデルの推定フェーズにおいて、DAG であることを保証する非巡回制約を用いた制約付きの連続最適化問題を解くことで DAG の推定を行う。その際、制約付き連続最適化問題を解くために、拡張ラグランジュ関数法を用いた反復的な部分問題の最適化を必要とする。そのため、潜在因子やサンプルサイズ、推定する辺の数の増加に伴い、計算負荷が増大するという欠点がある。

そこで、本研究では、この構造モデルの推定フェーズに着目し、DAG-NoCurl に基づくより効率的な推定アルゴリズムを提案した。DAG-NoCurl は、DAG を規定する重み付き隣接行列を新たに定式化することで、非巡回制約を必要としない制約なし連続最適化問題によって DAG を推定する手法である。DAG-NoCurl は、拡張ラグランジュ関数法を用いて制約付き最適化問題を解く LiNA と異なり、反復的な部分問題の最適化を必要としない。よって、DAG-NoCurl に基づく潜在因子間の因果探索アルゴリズムである提案手法は、既存手法である LiNA と比べて、より効率的な因果構造の推定が期待できる。

提案手法の有用性を確かめるために、本研究では、サンプルサイズ、潜在因子数、辺の数といった 3 つの要因を複数の水準に分けて数値実験を行った。その結果、提案手法は、既存手法と同等の推定精度を保ちつつ、推定にかかる計算時間を大幅に短縮できることが示された。また、本研究では、提案手法を用いた実データによる適用例も示した。実データは、心理学分野における教師の燃え尽き症候群に関するデータを用いた。その結果、理論的に示されている因果構造といくつかの部分で一致する構造を提示していることが確かめられた。(行動統計科学)