

潜在的交絡を考慮した因果探索（修論の続き）

長瀬 永遠

富山県立大学 情報基盤工学講座

January 20, 2023

卒論時点で抱えていた課題

因果探索に関するものだけを取り上げると以下の3つが挙げられる.

- データの時系列性が考慮されていない
- 未観測なデータへの考慮の不足
- データの巡回性の検討の必要性

VAR-LiNGAM

あるデータ群に属する項目 A の現在の値が同じデータ群の中の他の項目の過去の値からどのような影響を受けているかを分析することができるベクトル自己回帰モデル（VAR モデル）と通常の LiNGAM を組み合わせることで、現在のデータとの因果関係と過去のデータとの因果関係を同列に分析することができるモデル。

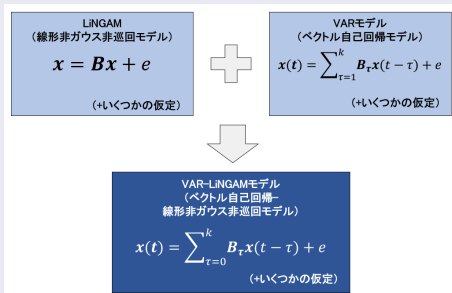


図 1: VAR-LiNGAM

未観測要因に対するアプローチ 1

4/7

なぜ、未観測要因を考慮する必要があるのか

一部の変数（因果関係を含む対のデータの一方）のみを観測した状態で分析を行うと誤った因果が導かれる可能性がある。



図 2: 未観測要因の影響

一方で、政策立案の分野において全てのデータを観測することは実質不可能。

Parce-LiNGAM

Parce-LiNGAM とは、未観測変数が存在することを考慮したうえでより確実に観測変数内の因果関係を分析するように LiNGAM を拡張したモデル。

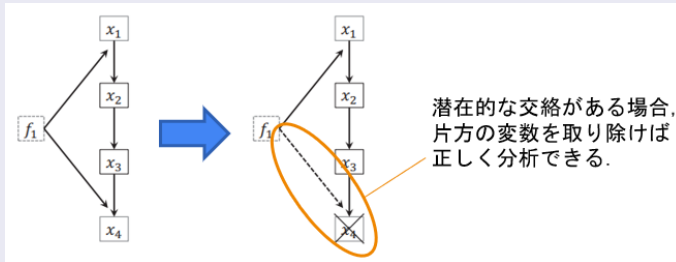


図 3: Parce-LiNGAM の基本アイデア

Parce-LiNGAM の基本的なアルゴリズム

ベースにあるのは観測変数の集合全体に対してではなく部分集合に分けて分析したのちに結果をマージするという考え方.

- 1 観測変数の集合を部分集合に分ける
- 2 各部分集合に対して LiNGAM を適用する
- 3 2 の結果に対してそれぞれ検定を行う
- 4 3 の結果をもとにマージを行う

ただし、上記の手順では計算量が膨大になるため、実際の Parce-LiNGAM では一度全体で分析を行った後に一部に対して再度分析を行うらしい. この辺のロジックはいまいち理解できていないので勉強中.

今後やること

- Parce-LiNGAM のロジックを理解し，Python で実装する
- 実際のデータに適用して結果を確認する
- VAR-LiNGAM と Parce-LiNGAM を組み合わせる