Use of Prior Knowledge to Discover Causal Additive Models with Unobserved Variables and its Application to Time Series Data

Takeshi Nicholas Maeda, Shohei Shimizu

蒲田 涼馬 (Ryoma Gamada)
u455007@st.pu-toyama.ac.jp

富山県立大学 電子・情報工学専攻 情報基盤工学部門

November 11, 2024

本研究の背景

はじめに
おわりに

背景

経済の発展や衰退の裏には様々な要素が絡み合っており，それらの相互関係や因果関係を推定することは経済の動向を知ったり，新たな金融政策を考える上で非常に重要なものである．しかし，経済における因果探索についての研究はそこまで充実しておらず，また未観測変数を考慮したリアルデータに直接適用できるシステムを開発している研究はさらに少ない．

本研究の目的

未観測変数を考慮した因果加法モデル (CAM-UV) の新たなふたつの手法を提案する．
1 つめの手法では効率的な因果発見のために事前知識を活用する手法，もう 1 つの手法では時系列データにおける因果関係を探索できる手法の提案を行う．
また，その有効性を証明するためにシミュレーションデータと実世界のデータの両方を使用する．時系列の方では既存の時系列因果発見手法と比較することで提案手法のテストを行う．

従来研究

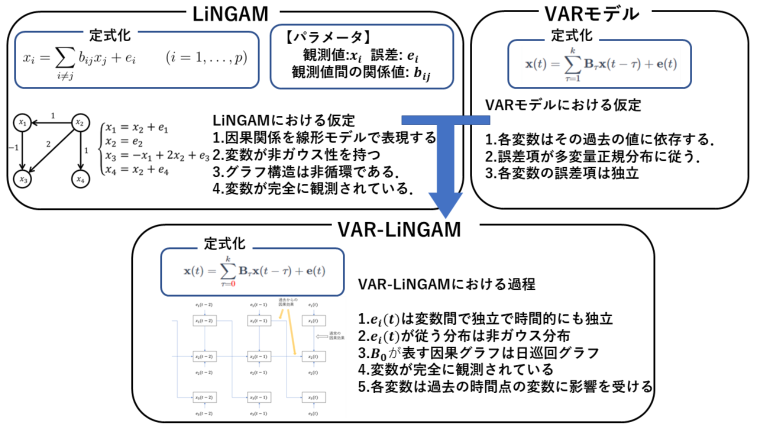
従来研究

因果関係発見とは因果関係を推論する統計的手法や機械学習手法のことを指す.
これらの研究では，データ作成過程に関する仮定から導かれる推論手法が提案されており，その手法により観測された変数間の因果グラフを作成することができる．その手法の 1 つとして因果構造とそれらの間の因果強度を求めることができる LiNGAM という手法がある.
しかしこの手法ではすべての変数が観測済みであるという仮定である．しかし，ほとんどのデータはそのような仮定を満たしていない.
潜在交絡因子の存在を仮定する手法の研究としては本研究の前段階の CAM-UV の研究などがある.
リアルデータに活用できるように本研究では CAM-UV の延長として研究を行っていく.

因果探索の手法1

LiNGAM

LiNGAM は統計的因果探索手法の 1 つである.



因果探索の手法4

CAM-UV

CAM-UV も同じく因果探索の手法であり，LiNGAM の拡張版である．CAM-UV と LiNGAM との違いはその仮定にある．CAM-UV では未観測変数を考慮しているのに加えて非線形のものも扱えるというメリットがある．
RCD と同じく未観測変数については双方向矢印 (←→) で表現する．
定義式は以下のようになる．

$$V_i = \sum_{Xj \in opa(V_i)} f_{i,j}(X_j) + \sum_{u_j \in upa(U_i)} f_{i,j}(U_j) + N_i \tag{1}$$

ここで $V = \{V_i\}$ は観察された変数または観察されていない変数の集合，$X = \{X_i\}$ は観察された変数の集合，$U = \{U_i\}$ は観察されていない変数の集合，N_i は V_i への外部効果，$opa(V_i) \subset X$ は V_i の観測されている直接原因の集合，$upa(V_i) \subset U$ は V_i の観測されていない直接原因の集合 $f_{i,j}$ は非線形関数である．

CAM-UV-PK

CAM-UV with prior knowledge(CAM-UV-PK)

この手法では X_i は X_j の原因にはなりえないといった文の形で事前知識を導入することができる．
これによって因果関係を探索する際の計算負担が減り，誤った因果関係の推定を抑制する効果がある．
基本式は以下の CAM-UV の基本式に変数として順序付きリストの T を加えたものになる．順序変数対 (X_i, X_j) が T に含まれる場合，それは X_i が X_j の直接または関節の原因にはなりえないと仮定されることを意味する．

$$V_i = \sum_{Xj \in opa(V_i)} f_{i,j}(X_j) + \sum_{u_j \in upa(U_i)} f_{i,j}(U_j) + N_i \tag{2}$$

TS-CAM-UV

TS-CAM-UV

TS-CAM-UV も CAM-UV の基本式と基本的に同じではあるが含まれる変数に違いがある．

$$X_{i,t} = f_i(X_{1,t-1}, X_{2,t-2}, \ldots, U_t) + \epsilon_{i,t} \tag{3}$$

変数に遅延性を含むことで時系列の分析に活用することができるようになり，因果グラフの構造にも (t-1) などの要素も含まれることになる．

CAM-UV-PK のデモ実験

CAM-UV-PK のデモ実験の結果

各サンプルサイズ

$$n \subset 100, 200, \cdots, 1900, 2000 \tag{4}$$

の人口データを用いて 100 回の実験を行い，本手法の既存の手法と比較した．因果効果は以下のように設定される．

$$\left(\sin\left(a_1\left(V_j + b_1\right)\right)\right)^3 c_1 + \left(\frac{1}{1 + \exp\left(-a_2\left(V_j + b_2\right)\right)} - 0.5\right) c_2 \cdots (11)$$

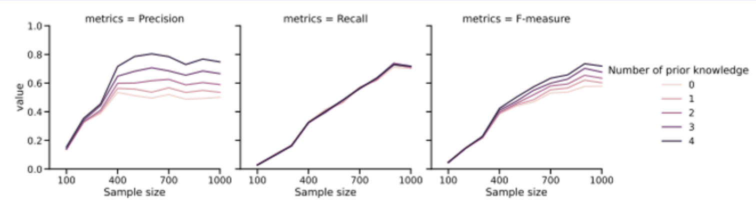


図 1: 左から精度再現率，F 値

TS-CAM-UV のデモ実験

はじめに
おわりに

CAM-UV-PK のデモ実験の結果

各サンプルサイズ

$$n \subset 100, 200, \cdots, 1900, 2000 \tag{5}$$

の人口データを用いて 100 回の実験を行い，本手法の既存の手法と比較した．因果効果は以下のように設定される．

$$\left(\sin\left(a_1\left(V_j + b_1\right)\right)\right)^3 c_1 + \left(\frac{1}{1 + \exp\left(-a_2\left(V_j + b_2\right)\right)} - 0.5\right) c_2 \cdots (11)$$

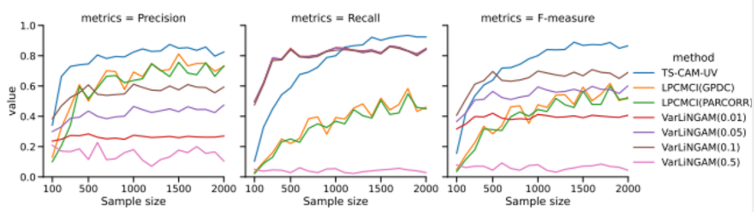
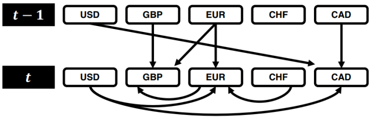


図 2: 左から精度再現率，F 値

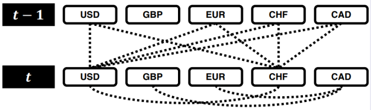
CAM-UV のリアルデータへの適用

CAM-UV の数値実験概要

時系列を考慮した CAM-UV(TS-CAM-UV) についての論文 (清水ら) を見つけたので，その論文と同じデータを用いて数値実験を行い比較してみた.
データはみずほ銀行によって公開されている日本円の公示外国為替データ (2021/10/26〜2023/11/8) までのものを用いている.



(a) The causal graph only with directed edges generated from TS-CAM-UV.



(b) The causal graph only with edges other than directed edges generated from TS-CAM-UV.

図 3: TS-CAM-UV についての論文参照

まとめ

はじめに
おわりに

まとめ

本稿では，CAM-UV の拡張として，CAM-UV-PK と TS-CAM-UV のふたつの手法を提案した．
CAM-UV-PK アルゴリズムはある変数が他のある変数の原因ではないという形で事前知識を導入する手法を採用する．
今後の研究では，同時期効果に周期があることを許容する因果関係発見法を探る必要があると考えられる．