

## 科学研究費助成事業 研究成果報告書

平成 28 年 5 月 9 日現在

機関番号：14401

研究種目：若手研究(B)

研究期間：2012～2015

課題番号：24700275

研究課題名(和文) 複数データセットからの高次元因果ネットワーク推定法の開発と生命科学への応用

研究課題名(英文) Estimation of high-dimensional causal networks based on multiple datasets and its applications to biomedical science

研究代表者

清水 昌平 (Shimizu, Shohei)

大阪大学・産業科学研究所・准教授

研究者番号：10509871

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：因果探索のためのモデルであるLiNGAMモデルにおいて複数データセットを融合する場合の推定原理の確立と、それに基づく直接推定法の開発を行った。また、未観測交絡変数がある場合へと拡張した。ヘテロ集団に対応するために未観測クラスがある場合も研究した。この拡張は、非線形モデルへの拡張とみなすこともできる。また、時間的に因果構造が変化するような場合へも拡張した。生命科学への応用としては、脳活動計測データの模擬データや実データに適用し、実用上の課題を見つけるとともに、分析から得られた情報を神経科学の共同研究者にフィードバックした。

研究成果の概要(英文)： We developed an estimation principle and estimation algorithms to estimate LiNGAM models from multiple datasets for causal structure learning. Moreover, we extended them to cases with latent confounding variables, latent classes, temporally changing causal structures. The latent class extension can be seen as a nonlinear extension. We applied those methods to brain imaging data.

研究分野：統計科学

キーワード：因果探索 非ガウス性 構造方程式モデル

## 1. 研究開始当初の背景

遺伝子間や脳領域間の因果構造の解明は、病気の理解や治療法の開発に非常に役立つ。例えば、がんは遺伝子ネットワークの異常、アルツハイマーは脳領域をつなぐ神経細胞ネットワークの異常が原因の1つと言われている。そのため、マイクロアレイによる遺伝子発現データや脳機能イメージングによる脳活動データを用いて、因果ネットワークを推定するための統計解析法が望まれている。例えば、「遺伝子 A を発現させると、遺伝子 B が発現する」といった関係を推定したい。因果ネットワーク推定の従来法はベイジアンネットワークである。しかし、ベイジアンネットワークには識別性の問題があり、データに同等に適合するモデルが複数存在する。そのため、たとえモデルの仮定が正しくても、多くの因果ネットワークを原理的に同定できない。さらに、推定法も局所解の問題があり、初期値を変えて多数回実行しても、良い解を得られないことは頻繁にある。

そこで申請者が考案したのが、LiNGAM モデルによるアプローチである。LiNGAM モデルは、構造方程式モデル(Bollen, 1989)で外的影響変数に非ガウス分布を仮定したモデルである。この非ガウス性の仮定によって、従来法が識別できない因果ネットワークの多くを一意に同定できる。モデル推定には、反復法ではなく直接法を開発しており、初期値選択をする必要がない。

応用としては、脳領域ネットワーク推定(Smith et al., 2011, NeuroImage)だけでなく、社会科学・環境科学などへも広く適用されている。

データの非ガウス性を活用することで、従来法では不可能だった分析が可能になる場合があることがわかってきた。ただ、実際の適用場面では十分なサンプルサイズの確保が難しい場合がある。例えば、マイクロアレイデータの場合は、コストなどの問題によってサンプルサイズが限られてしまう。ただし、一つ一つのデータセットのサンプルサイズが十分でなくても、同じ目的の下で異なる実験条件のデータセットが得られていることは多い。そこで、そのような複数のデータセットを融合して擬似的にサンプルサイズを増やし、推定精度を向上させることを着想した。実験条件が異なるので、もちろん完全に同じではない。だが、同じ目的の下で取得されたデータなので、なんらかの類似性はある。その類似性を利用し、推定精度を上げるのである。

## 2. 研究の目的

遺伝子発現データや脳機能イメージングデータのような高次元データから、因果ネットワークを高い精度で推定するための統計解

析法を開発する。

これまで多くの推定法が提案されてきたが、実際の適用場面では十分なサンプルサイズの確保が難しく、推定精度が低くなりがちである。しかし、異なる条件下で複数のデータセットを取得することは比較的容易である。そこで本研究では、複数のデータセットを効果的に融合させることで推定精度を高めることを目指す。

## 3. 研究の方法

本研究は4年間で、(1) 複数データセットからのLiNGAMモデル推定法の開発、(2) LiNGAMモデルの拡張、(3) 実データによる検証・実問題への適用を行う。

以下、もう少し細かい研究項目を挙げる。

複数データセットを融合するための推定原理の確立と

その原理に基づく直接推定法の開発を行う。

LiNGAMモデルを潜在変数・時間構造・非線形関係がある場合に拡張し、方法論の適用範囲を大きく広げる。未観測の遺伝子や脳領域などの潜在変数がある場合にも、因果に関する有用な情報を取り出したり、遺伝子制御や脳活動などの時間構造を利用したり、非線形関係を推定できるようにしたりする。

公開されている実データを用いて方法論の妥当性を検証する。実質科学の背景を考慮した上で結果を解釈し、方法論研究・応用研究にフィードバックする。

## 4. 研究成果

LiNGAMモデルにおいて複数データセットを融合する場合の推定原理の確立と、それに基づく直接推定法の開発を行った。

を未観測交絡変数がある場合へと拡張した。具体的には、LiNGAMモデルを未観測交絡変数がある場合に拡張し、Latent variable LiNGAMモデル上で因果分析法について研究をした。因果構造の探索と因果構造の探索と未観測交絡変数の検出を同時に切り分けて、未観測交絡変数を無視するのではなく、未観測交絡変数の影響のない(小さい)部分と大きい部分とを切り分けて、未観測交絡変数に対してより頑健な方法を研究した。複数データセット利用のフレームワークと組み合わせ、未観測交絡変数がある場合についても、複数データセットを融合することで推定精度が大幅に向上することを数値実験で確認した。

LiNGAMによる因果構造推定結果の信頼性を評価するために、LiNGAMモデルにおけるブートストラップ法による信頼区間の構成法についての研究も行った。

ヘテロ集団に対応するために未観測クラスがある場合も研究した。この拡張は、非線形モデルへの拡張とみなすこともできる

時間的に因果構造が変化するような場合に対応するため、複数データセット融合のアイデアを拡張し、経時データにおける因果構造推定法を考案した。

生命科学への応用としては、脳活動計測データの模擬データや実データに適用し、実用上の課題を見つけるとともに、分析から得られた情報を神経科学の共同研究者にフィードバックした。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 4 件)

(1) Shohei Shimizu and Kenneth Bollen. Bayesian estimation of causal direction in acyclic structural equation models with individual-specific confounder variables and non-Gaussian distributions. *Journal of Machine Learning Research*, 15: 2629-2652, 2014. 査読有  
URL:  
<http://jmlr.org/papers/volume15/shimizu14a/shimizu14a.pdf>

(2) Tatsuya Tashiro, Shohei Shimizu, Aapo Hyvarinen, and Takashi Washio. ParceLiNGAM: A causal ordering method robust against latent confounders. *Neural Computation*, 26(1): 57-83, 2014. 査読有  
DOI:10.1162/NECO\_a\_00533

(3) Shohei Shimizu. LiNGAM: Non-Gaussian methods for estimating causal structures. *Behaviormetrika*, 41(1): 65-98, 2014. 査読有  
DOI: 10.2333/bhmk.41.65

(4) Shohei Shimizu. Joint estimation of linear non-Gaussian acyclic models. *Neurocomputing*, 81: 104-107, 2012. 査読有  
DOI:10.1016/j.neucom.2011.11.005

[学会発表](計 26 件)

(1) Shohei Shimizu. (2016 年 3 月 23 日). Non-Gaussian structural equation models for causal discovery. 2016 Probabilistic Graphical Model Workshop: Sparsity, Structure and High-dimensionality, 統計数理研究所(東京). 招待あり

(2) 清水昌平. (2016 年 3 月 17 日). 因果探索: 基本から最近の発展までを概説. 第 23 回情報論的学習理論と機械学習研究会 (IBISML), 統計数理研究所(東京). 招待あり

(3) 清水昌平. (2016 年 3 月 16 日). 因果探索: 観察データから因果仮説を探索する. 日本社会心理学会: 第 3 回春の方法論セミナー, 上智大学(東京). 招待あり

(4) 清水昌平. (2016 年 3 月 1 日). 因果探索と非ガウス性, 第 11 回協定講座シンポジウム: 計算科学とビジュアル・アナリティクス, 神戸大学(兵庫). 招待あり

(5) Shohei Shimizu. (2016 年 2 月 5 日). Statistical estimation of causal directions based on observational data, The 3rd CiNet Conference - Neural Mechanism of Decision Making: Achievements and New Directions, 情報通信研究機構 脳情報通信融合研究センター(大阪). 招待あり

(6) Shohei Shimizu. (2016 年 1 月 7 日). Non-Gaussian methods for causal discovery. International Workshop on Causal Inference, 統計数理研究所(東京). 招待あり

(7) Shohei Shimizu. (2015 年 11 月 17 日). A non-Gaussian approach for causal discovery in the presence of hidden common causes. Second Workshop on Advanced Methodologies for Bayesian Networks (AMBN2015), 慶応義塾大学(神奈川). 招待あり

(8) Patrick Blobaum, Shohei Shimizu, Takashi Washio. (2015 年 11 月 17 日). Discriminative and generative models in causal and anticausal settings. Second Workshop on Advanced Methodologies for Bayesian Networks (AMBN2015), 慶応義塾大学(神奈川).

(9) 清水昌平. (2015 年 9 月 2 日). 因果探索: データから因果の方向性等を調べる. 日本行動計量学会 第 43 回大会 ラウンドテーブル: 徹底討論・統計的因果推論 データだけから因果を言えるのか? 3 つのアプローチから. 首都大学東京(東京). 招待あり

(10) 清水昌平. (2015 年 8 月 6 日). 非ガウス性を利用した因果構造探索. 2015 年日本生態学会関東地区会シンポジウム「非ガウス性 / 非線形性 / 非対称性からの

因果推論手法：その使いどころ・原理・実装を学ぶ」,  
東京大学(東京). 招待あり

(11) 清水昌平. (2015年3月20日). 構造方程式モデルによる因果探索と非ガウス性. 数学協働プログラム ワークショップ: 確率的グラフィカルモデル, 電気通信大学(東京). 招待あり

(12) Shohei Shimizu. (2014年6月25日). Estimation of causal direction in the presence of latent confounders and linear non-Gaussian structural equation models. Causal Modeling and Machine Learning (Post-ICML workshop), Beijing(China). 招待有

(13) Shohei Shimizu. (2014年5月23日). A non-Gaussian approach for estimating possible causal direction in the presence of latent confounders. Conference on Statistics and Causality 2014, Vienna(Austria). 招待あり

(14) 清水昌平, Kenneth Bollen  
Estimation of causal direction in the presence of latent confounders using LiNGAM  
第16回情報論的学習理論ワークショップ (IBIS2013)  
2013年11月12日  
東京工業大学 (東京)

(15) Kento Kadowaki, Shohei Shimizu, Takashi Washio  
Estimation of causal structures in longitudinal data using non-Gaussianity  
23rd IEEE International Workshop on Machine Learning for Signal Processing (MLSP2013)  
2013年9月25日  
Southampton (United Kingdom)

(16) 清水昌平  
LiNGAMによる因果構造推定: 潜在交絡変数がある場合  
科研費シンポジウム: Incomplete Data Analysis and Causal Inference  
2013年9月23日  
大阪大学 (大阪)

(17) Naoki Tanaka, Shohei Shimizu, Takashi Washio  
Estimation of causal direction in the presence of latent confounders using a Bayesian LiNGAM mixture model  
Causality: Perspectives from Different Disciplines  
2013年8月8日

Vals (Switzerland)

(18) 門脇健人, 清水昌平, 鷺尾隆  
経時データにおける非ガウス性を用いた因果構造探索  
第27回人工知能学会全国大会  
2013年6月6日  
富山国際会議場 (富山)

(19) 田中直樹, 清水昌平, 鷺尾 隆  
潜在交絡変数が存在する場合のベイズ的アプローチによる非ガウス因果構造推定法  
第27回人工知能学会全国大会  
2013年6月6日  
富山国際会議場 (富山)

(20) Kittitat Thamvitayakul, Shohei Shimizu, Tsuyoshi Ueno, Takashi Washio, Tatsuya Tashiro  
Bootstrap confidence intervals in DirectLiNGAM  
2012年12月10日  
2012 IEEE 12th International Conference on Data Mining Workshops (ICDMW2012)  
Brussels (Belgium)

(21) 清水昌平  
因果構造探索と非ガウス構造方程式モデル.  
2012年11月17日  
人工知能学会 第87回人工知能基本問題研究会 (SIG-FPAI) (招待講演)  
慶應義塾大学 (神奈川)

(22) 田代竜也, 清水昌平, Aapo Hyvärinen, 鷺尾 隆  
非ガウス構造方程式モデルにおける因果順序の推定: 潜在交絡変数に頑健な方法  
2012年9月24日  
第3回 Latent Dynamics ワークショップ  
東京大学 (東京)

(23) Tatsuya Tashiro, Shohei Shimizu, Aapo Hyvärinen, Takashi Washio  
Estimation of causal orders in a linear non-Gaussian acyclic model: a method robust against latent confounders. 22nd Int. Conf. on Artificial Neural Networks (ICANN2012)  
2012年9月13日  
Lausanne (Switzerland)

(24) 清水昌平.  
構造方程式モデルによる因果推論: 因果構造探索に関する最近の発展  
日本行動計量学会第40回大会 (招待講演)  
2012年9月13日  
新潟県立大学 (新潟)  
チュートリアル

(25) Kittitat Thamvitayakul, 清水昌平,

鷲尾 隆, 田代竜也  
Bootstrapping confidence intervals in  
linear non-Gaussian causal model  
第 26 回人工知能学会全国大会  
2012 年 6 月 15 日.  
山口県教育会館 (山口)

(26) 田代竜也, 清水昌平, 鷲尾 隆.  
非ガウス性を用いた線形非巡回なデータ生  
成過程部分の発見と同定.  
第 26 回人工知能学会全国大会  
2012 年 6 月 15 日  
山口県教育会館 (山口)

〔図書〕(計 0 件)

〔産業財産権〕  
出願状況(計 0 件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
国内外の別：

取得状況(計 0 件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
取得年月日：  
国内外の別：

〔その他〕  
ホームページ等  
<https://sites.google.com/site/sshimizu06/indexj>

## 6. 研究組織

(1)研究代表者  
清水 昌平 (SHIMIZU, Shohei)  
大阪大学・産業科学研究所・准教授  
研究者番号：10509871

(2)研究分担者  
該当なし

(3)連携研究者  
該当なし