

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

田代 竜也* 清水 昌平† Aapo Hyvärinen ‡ 鷲尾 隆§
Tatsuya Tashiro Shohei Shimizu Takashi Washio

Abstract: 近年様々な分野で大量の観測データが蓄積されており、因果分析法に対するニーズは高まっている。最近の研究により、データの非ガウス性を利用することで変数間の因果的順序を同定できる場合があることがわかっている。本研究では因果順序の推定と同時にモデルがデータに適合しているかを検定する手法を提案する。これにより、未観測交絡変数がある場合に頑健な推定することが可能となる。

Keywords: 因果分析, 非ガウス, 潜在交絡変数

1 研究概要

本研究では、対象とする事象の観測データから各変数間の因果関係を推定し、グラフィカルに表現することを目的とする。因果分析の研究分野はバイオインフォマティクスやニューロンインフォマティクス等さまざまな分野への応用が期待できる。因果分析においては構造方程式モデル (Structural Equation Model, SEM) が広く用いられてきた [1]。しかし、構造方程式モデルは一般的に線形ガウス性を仮定しており、因果構造に対する事前情報なしではモデルを一意的に同定することができないという識別性の問題を有している。また、現実にはガウス分布に従わないデータも多数存在する。そこで近年、非ガウス性を仮定することにより、事前情報を用いずに観測データのみから線形非巡回モデルの同定が可能であることが示された。このモデルは LiNGAM (Linear Non-Gaussian Acyclic Model) モデル [2] と呼ばれ、独立成分分析 (Independent Component Analysis, ICA) [3] と密接に関連している。

LiNGAM モデルを推定する手法には、ICA を基にした ICA-LiNGAM アルゴリズム [2] と、ICA を使用せずダイレクトに因果的順序を上から順に推定する DirectLiNGAM アルゴリズム [4] が提案されている。しかしながら、これらの手法は LiNGAM モデルの全ての仮定が満たされていることを前提として因果構造を分析するため、いずれかの仮定が満たされていない場合、完全に誤った推定結果を出力することがある。LiNGAM モデルの仮定が破綻する典型的な原因の一つが潜在交絡変数である。潜在交絡変数とは、観測変数に因果的影響を与えるような未観測の変数である。そこで、潜在交絡変数を考慮したモデルである LvLiNGAM (Latent variable LiNGAM) モデル [5] が提案されている。LvLiNGAM モデルを推定する手法として、過完備独立成分分析 (Overcomplete ICA) を基にした LvLiNGAM アルゴリズム [5] と、潜在交絡変数からの影響を受けていない 2 変数を探索し、その 2 変数間の順序を推定する Pairwise LvLiNGAM アルゴリズム [6] が提案されている。しかし、LvLiNGAM アルゴリズムは過完備独立成分分析の計算負荷が高いため、4 変数以上のデータへの適用が困難である。また、Pairwise LvLiNGAM アルゴリズムはその性質上、3 変数以上の因果的順序は推定されない。そこで本発表では DirectLiNGAM アルゴリズムを拡張し、潜在交絡変数から直接影響を受けていない変数間の因果構造を可能な限り推定する新たな手法について述べる。この手法は 5, 6 変数以上のデータに対しても十分に適用でき、適用するデータによっては 3 変数以上の間の因果的順序を推定することが可能である。尚、本発表は [7] に準拠している。

*大阪大学 産業科学研究所, 567-0047 大阪府茨木市美穂ヶ丘 8-1, e-mail tashiro@ar.sanken.osaka-u.ac.jp,

The Institute of Scientific and Industrial Research, Osaka University, 8-1, Mihogaoka, Ibarakishi, Osaka, 567-0047, Japan

†大阪大学 産業科学研究所, 567-0047 大阪府茨木市美穂ヶ丘 8-1, e-mail sshimizu@ar.sanken.osaka-u.ac.jp

The Institute of Scientific and Industrial Research, Osaka University 8-1, Mihogaoka, Ibarakishi, Osaka, 567-0047, Japan

‡Dept. of Computer Science and Dept. of Mathematics and Statistics, University of Helsinki, Helsinki Institute for Information Technology, FIN-00014, Finland e-mail aapo.hyvarinen@helsinki.fi

§大阪大学 産業科学研究所, 567-0047 大阪府茨木市美穂ヶ丘 8-1, e-mail washio@ar.sanken.osaka-u.ac.jp

The Institute of Scientific and Industrial Research, Osaka University 8-1, Mihogaoka, Ibarakishi, Osaka, 567-0047, Japan

参考文献

- [1] K. Bollen. *Structural Equations with Latent Variables*. John Wiley & Sons, 1989.
- [2] S. Shimizu, P. O. Hoyer, A. Hyvärinen, and A. Kerminen. A linear non-gaussian acyclic model for causal discovery. *J. Mach. Learn. Res.*, Vol. 7, pp. 2003–2030, 2006.
- [3] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent component analysis*. Wiley, New York, 2001.
- [4] S. Shimizu, T. Inazumi, Y. Sogawa, A. Hyvärinen, Y. Kawahara, T. Washio, P. O. Hoyer, and K. Bollen. DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model. *J. Mach. Learn. Res.*, Vol. 12, pp. 1225–1248, 2011.
- [5] P. O. Hoyer, S. Shimizu, A. Kerminen, and M. Palviainen. Estimation of causal effects using linear non-gaussian causal models with hidden variables. *International Journal of Approximate Reasoning*, Vol. 49, No. 2, pp. 362–378, 2008.
- [6] D. Entner and P. O. Hoyer. Discovering unconfounded causal relationships using linear non-gaussian models. In *New Frontiers in Artificial Intelligence, Lecture Notes in Computer Science*, Vol. 6797, pp. 181–195, 2011.
- [7] T. Tashiro, S. Shimizu, A. Hyvärinen, and T. Washio. Estimation of causal orders in a linear non-gaussian acyclic model: a method robust against latent confounders. In *Proc. Int. Conf. on Artificial Neural Networks (ICANN2012)*, Lausanne, Switzerland, 2012.