

時系列間の影響波及の遅れを考慮した縮小推定による因果指標の拡張

山下 裕也†

渋谷 崇†

原田 達也†

國吉 康夫†

† 東京大学大学院情報理工学系研究科

1 はじめに

経済，脳科学などの分野において，予測や構造推定の為に時系列データ間の因果関係を推定する事が行われている．ここで言う因果関係は，ある時系列の過去の変動が別の時系列の変動に影響を与えている時の関係を指す．これまで因果関係の強さを定量化する為のいくつかの手法が提案されてきているが，従来の手法ではある時系列の直近の変動が他の時系列に影響を与えていると仮定しており，影響の伝搬に時間差がある場合を考慮していない．この為，過去のどの時点から影響を受けているかが分からず，予測を行う場合も過去の不必要な情報をモデルに組み込んでしまう場合がある．

そこで本研究ではモデルの推定に縮小推定の一種である最小角回帰法を利用する事で，ある時系列から他の時系列への影響の伝わる時間遅れを推定できる様に従来の因果指標を拡張する．また人工データを用いた実験によってその性能の評価を行う．

2 Granger Causality

これまで二つの連続値時系列間の因果性を定量化する手法がいくつか提案されてきている．その中でもGranger Causality (GC) [1] は経済学・脳科学などで広く利用されており，本研究でもGCを元に拡張を行う．GCでは時系列 y から x の因果性を見るときに次の二つの時系列モデルを比較する．

$$x_t = \mathbf{a}^T \mathbf{x}_{t-1}^{(m)} + \varepsilon_{x|x,t}, \quad (1)$$

$$x_t = \mathbf{b}^T \mathbf{x}_{t-1}^{(m)} + \mathbf{c}^T \mathbf{y}_{t-1}^{(m)} + \varepsilon_{x|xy,t}, \quad (2)$$

ここで， \mathbf{a} ， \mathbf{b} ， \mathbf{c} は係数， x_t は時刻 t における x の観測値， $\mathbf{x}_{t-1}^{(m)}$ は時刻 $t-1$ から $t-m$ の観測値を並べたものであり， y についても同様である．また， $\varepsilon_{x|x,t}$ ， $\varepsilon_{x|xy,t}$ はそれぞれ平均0，分散 $\sigma_{x|x}^2$ ， $\sigma_{x|xy}^2$ の誤差項である．この場合の因果関係の強さは次の様に表現される [2]．

$$G_{y \rightarrow x} = \frac{1}{2} \log_2 \frac{\sigma_{x|x}^2}{\sigma_{x|xy}^2}. \quad (3)$$

GCでは y の過去の値が x の未来の値の予測精度を良くするならば $\sigma_{x|xy}^2$ が $\sigma_{x|x}^2$ より小さくなり，その改善の度合いを因果の強さだとみなす．

3 縮小推定による Granger Causality の拡張

式 (1) の様な自己回帰モデルでは，情報量規準などによって最適な m を決定するといった事が行われている．この場合 $t-1$ から $t-m$ がモデルに組み込まれているが，直近の時刻 $t-1$ などから常に影響があるとは限らず，データによっては遅れて影響が伝わる事もありうる．そこで縮小推定の一種である最小角回帰法 (Least Angle Regression, LARS) [3] を組み合わせ変数選択を行う事によって，原因から結果への影響波及の時間遅れを考慮したGCの拡張を行う．

LARSは回帰モデルの推定の際に目的変数と相関の高い説明変数から再帰的にモデルに組み込む手法である．例として式 (1) の場合，各段階で選択されている変数のみから予測した時刻 t での x の値を \hat{x}_t とすると，LARSでは $\mathbf{x}_{t-1}^{(m)}$ の各要素の内， $x_t - \hat{x}_t$ との相関が最大となるものをモデルに加えていき，最終的に全ての説明変数がモデルに組み込まれる．LARSの途中の結果から最良のモデルを選択する事によって，原因を表す時系列から結果を表す時系列への影響波及の遅れを推定する事が可能になる．本研究では情報量規準を用いて，このモデル選択を行う．

GCの計算においてLARSを使う事の問題として，そのままLARSを式 (1,2) のモデル推定に適用した場合，式 (2) の誤差の分散 $\sigma_{x|xy}^2$ が式 (1) の誤差の分散 $\sigma_{x|x}^2$ よりも小さくなるとは限らない事が挙げられる．そこで，指標の値が0以上になる様に次のモデルの比較によって因果の強さを計算する事を提案する．

$$x_t = \mathbf{a}^T \mathbf{x}_{t-1}^{(m)} + \varepsilon_{x|x,t}, \quad (4)$$

$$\mathbf{y}_{t-1}^{(m)} = \mathbf{B}^T \hat{\mathbf{x}}_{t-1}^{(m)} + \varepsilon_{y|x,t-1}, \quad (5)$$

$$\varepsilon_{x|xy,t} = \mathbf{c}^T \varepsilon_{y|x,t-1} + \varepsilon_{x|xy,t}, \quad (6)$$

因果の強さは，式 (4) と式 (6) の誤差の分散を用いて，式 (3) から計算出来る．式 (5) では式 (4) で $\mathbf{x}_{t-1}^{(m)}$ から選択された要素を成分に持つ $\hat{\mathbf{x}}_{t-1}^{(m)}$ を用いる．ただし，式 (5) のみ情報量規準による変数選択は行わない．式 (6) から， $\varepsilon_{x|xy,t}$ の分散は必ず $\varepsilon_{x|x,t}$ の分散より小さくなる．

LARSは目的変数が一次元であるときの手法であるが，時系列データが多次元の場合でも提案手法は適用可能である．Ladroueら [4] は多次元時系列データでのGCにおいて，誤差のトレースによる比較でも十分な精度が得られる事を示している．これは誤差の各次元ご

Extension of causality measure using shrinkage estimation
†Y. YAMASHITA, †T. SHIBUYA, †T. HARADA, and †Y. KUNIYOSHI
†The University of Tokyo

との関係性を考慮しなくても十分な性能を得られるという事である．その為、回帰モデルの目的変数の各次元に対して一次元の場合と同様の最小角回帰法と情報量基準による変数選択を適用する事が可能であり、得られる誤差の分散共分散行列のトレースで評価出来る．

4 評価実験

実験では二種類の人工データを用いて時間遅れの推定、及び得られたモデルの予測性能について提案手法と従来の GC の比較を行った．どちらも次の様なモデルから生成される時系列 x, y を用いた．

$$\begin{cases} x_t = \varepsilon_t, \\ y_t = 0.5 y_{t-1} + \mathbf{a}^T \mathbf{x}_{t-1}^{(20)} + \varepsilon_t, \end{cases} \quad (7)$$

ε_t は平均 0、分散 0.1 のガウシアンノイズであり、このデータは $x \rightarrow y$ の方向にのみ影響を持つ．実験では二種類の \mathbf{a} を用いた．一つはある時系列から他方の時系列へ一定の時間差で影響が伝わる様なデータであり、もう一つは特定の時間差で影響が伝わるのではなく、直前を含む過去の一定の時間間隔から影響を受ける様なデータである．それぞれの場合の \mathbf{a} を図 1 にグラフで示す．グラフの絶対値が大きいほど、その時間差で伝わる影響が強くなる．実験ではそれぞれの \mathbf{a} に対応したモデル 1, 2 からサンプル数が 100 と 1000 のデータを生成した．サンプル数 100 のデータから因果の強さとモデルを推定し、推定したモデルでサンプル数 1000 のデータに対する予測性能を見る．また提案手法ではベイズ情報量規準を用いて変数選択を行った．

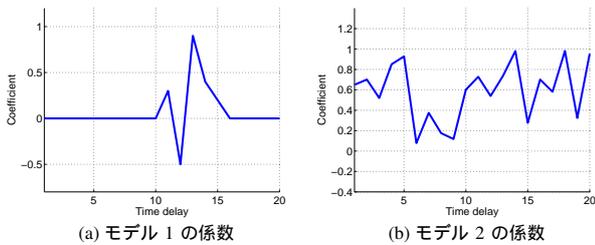


図 1: 人工データの \mathbf{a} の値

100 サンプルのデータに対する実験結果を図 2 に示す．推定した \mathbf{a} の値について、図 2(a) で提案手法では実際に影響を持っている $t-11$ から $t-15$ の係数が大きくなっており、その他の部分では $t-4$ をのぞき 0 になっている．それに対し従来の GC では全体的に係数の値が大きくなっており、特定の時刻から影響を受けている事は読み取れない．一方モデル 2 の場合はどの時間差も影響を持つ為に変数選択が必要無く、どちらの手法でもほぼ同じ係数が推定されている．ただし推定した係数のスケールが図 1 と異なるのは、一度データを正規化してから各手法を適用している為である．

また因果関係の強さについては、関係性の無い方向 ($y \rightarrow x$) に対する指標の値が GC に比べて提案手法の方

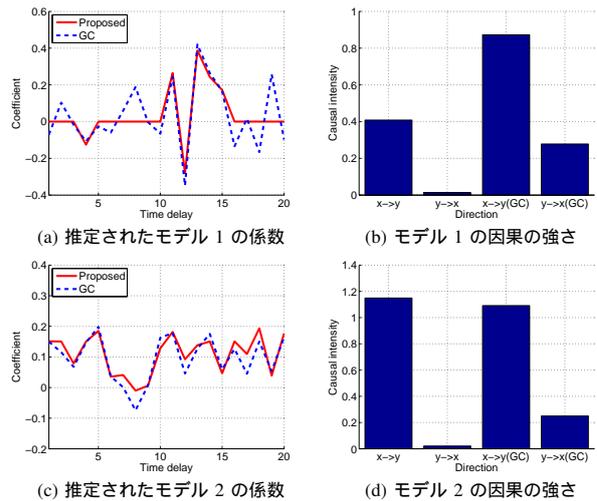


図 2: 実験結果. (a,c) 各時系列データから予測した $\mathbf{x}_{t-1}^{(20)}$ の係数．(b,d) 提案手法と GC で求めた各方向における因果の強さ．

が小さく抑えられている事が分かる．図 2(b) で、関係性がある方向の指標値も提案手法では小さくなっているが、関係性がない方向に対する比率としては提案手法の方が大きく、因果関係の有無を調べる点では問題無いと考えられる．

次に、100 サンプルで推定したモデルを用いて、同じ生成モデルから得られた別の 1000 サンプルの予測を行った場合の平均二乗誤差を表 1 に示す．提案手法では、時間遅れがある場合は予測に必要な変数を取り除けるので汎化性が GC に比べ高くなっている事が分かる．また、モデル 2 の様な場合でも性能が落ちていない事が確認できる．

表 1: 平均二乗誤差による予測性能の比較

Model1(Proposed)	Model1(GC)	Model2(Proposed)	Model2(GC)
0.4596	0.8455	0.2094	0.2212

5 まとめ

本研究では、縮小推定を用いて従来の因果指標を拡張する事により、ある時系列から別の時系列へ影響が伝わる場合の時間差の推定を可能にした．また提案手法で得られたモデルの予測性能が従来の手法より良くなる事も実験で示した．

参考文献

- [1] C. W. J. Granger. Investigating causal relations by econometric models and cross spectral methods. *Econometrica*, Vol. 37, No. 3, pp. 424-438, 1969.
- [2] T. Shibuya, T. Harada and Y. Kuniyoshi. Causality quantification and its applications: Structuring and modeling of multivariate time series. *ACM SIGKDD*, 2009.
- [3] B. Efron, T. Hastie, I. Johnstone and R. Tibshirani. Least Angle Regression. *Annals of Statistics*, Vol. 32, No. 2, pp. 407-499, 2004.
- [4] C. Ladrone, S. Guo, K. Kendrick and J. Feng. Beyond element-wise interactions: Identifying complex interactions in biological processes. *PLoS ONE*, Vol. 4, 2009.