

複雑システムの構造推定のためのグレンジャー因果性に基づく閾値決定法

上岡 祐太[†] 鈴木 智也^{††} 山本 誠一[‡][†]同志社大学大学院 ^{††}茨城大学工学部 [‡]同志社大学理工学部

1 はじめに

現実世界には多数の要素が相互作用して複雑な挙動を示すシステムが多く存在する．このような複雑システムを解析するためには，システムから観測可能な時系列データのみを用いて構造を推定することが重要である．システムの構造，つまり相互作用のネットワーク構造を知ることができれば，システムのモデル化や将来変動の予測を行う上で有利になる．このような観点から，システムの因果構造を推定する方法として様々な手法 [1]–[4] が提案されている．しかし，計算結果から要素間の結合の有無を判定するためには閾値を定める必要がある．この閾値は，経験的または恣意的に決定される場合が多い．そこで本研究では，グレンジャーの因果性 [5] に基づいてこの閾値を計算的に定める方法を提案する．

2 予測精度最大化による閾値決定法

各要素が生み出す時系列データ間の相関係数や偏相関係数を用いて要素間の相互作用の有無を分析することは従来より一般的に行われている．この推定値 $\tilde{W}_{i,j}$ ($i \neq j$) は，要素 i, j 間の相互作用の強さ，つまり結合の可能性を表している．しかし， $\tilde{W}_{i,j}$ は連続量であるため結合の有無を判別するためには，適切な閾値 θ を設定する必要がある．従来では，恣意的に θ を定める場合が多かったが，本研究ではグレンジャーの因果性 [5] を利用してこの θ を計算的に定める．

グレンジャーの因果性とは，予測精度を向上させる要素は全て，予測対象の要素に影響を与えていると考える．例えば，要素 i の将来変動を予測する場合，予測モデルから要素 j の効果を取り除いた時に予測精度が悪化するのであれば，要素 j は要素 i に影響を与えていると判断する．このような予測精度に基づいた考え方を応用すれば，前述の推定値 $\tilde{W}_{i,j}$ から結合の有無を判別する閾値 θ を定めることができる．

まずシステムを構成する要素 i ($i = 1, 2, \dots, N$) から時系列データ x_t^i を観測し，相関係数や偏相関係数を用いて推定値 $\tilde{W}_{i,j}$ を求める．この $\tilde{W}_{i,j}$ をもとに，式 (1)(2) で表される予測モデル [6] を構成する．この予測モデルは，非線形データに対しても短期的には予測が可能であり，また $\tilde{W}_{i,j}$ により相互作用の強さを

反映することができる．

$$X_t^i = \{\tilde{W}_{i,1}x_t^1, \dots, x_t^i, \dots, \tilde{W}_{i,N}x_t^N\} \quad (1)$$

$$\tilde{X}_{t+1}^i = \frac{1}{K} \sum_{k=1}^K X_{t_k+1}^i \quad (2)$$

ここで \tilde{X}_{t+1}^i は予測値であり， $X_{t_k}^i$ は X_t^i の近傍点である．本研究では $K = N + 1$ とした．なお， \tilde{X}_{t+1}^i 中に予測値 \tilde{x}_{t+1}^i が含まれており，真値 x_{t+1}^i との相関係数をこの要素 i の予測精度 Λ^i とした．

次に，この予測モデルに対して，ある閾値 θ を用意し， θ 以下の $\tilde{W}_{i,j}$ を 0 に設定する．つまり $\theta \geq \tilde{W}_{i,j}$ を満たす要素 j は予測対象 i と関連しない可能性が高いので，予測モデルから除去する．このとき式 (1)(2) は θ によって変化するため， X_t^i を $X_t^i(\theta)$ と表現すると，式 (2) は次式となる．

$$\tilde{X}_{t+1}^i(\theta) = \frac{1}{K} \sum_{k=1}^K X_{t_k+1}^i(\theta) \quad (3)$$

さらに，予測精度 Λ^i も θ に依存するので $\Lambda^i(\theta)$ と書く．この予測を全要素に対して行い， $\Lambda^i(\theta)$ の平均をとったものを最終的な予測精度 $\Lambda(\theta)$ とする．グレンジャーの因果性によれば，予測精度が高いほど要素間の因果構造を適切に予測モデルに取り込めることを支持できるので，式 (3) 中の θ を少しずつ変化させながら繰り返し予測を行い，予測精度 $\Lambda(\theta)$ が最大となる時の θ を結合の有無を判別する最適な閾値とする．

3 評価実験

本手法の有用性を確かめるために，人工的に作成した複雑システムに対して因果推定を行う．まず相互作用を表現する数理モデルとして，D. Watts と S. Strogatz による WS モデル [7] を用いた．WS モデルは張り替え確率 p により，様々な構造を実現できる． $p = 0$ の時は局所的な結合のみを持つレギュラーネットワークとなり， $p = 1$ の時は完全なランダムネットワークとなる．その中間である $0 < p < 1$ の時はスモールワールドネットワークの状態になる．さらにネットワーク上で駆動するダイナミクスとして，Coupled Map Lattice (CML) [8] を用いる．しかし通常の CML は局所的な相互作用しか持たないレギュラーネットワーク構造であるため，WS モデルを適用し式 (4) のように修正した．

$$x_{t+1}^i = f \left((1-w)x_t^i + \frac{w}{N_i} \sum_{j \in G_i(p)} x_t^j \right) \quad (4)$$

ここで， $G_i(p)$ は i 番目の要素と結合している要素 j の集合であり， N_i は $G_i(p)$ の要素数である． w は結

Threshold selection based on the Granger causality to identify the structure of complex systems.

[†] Yuta UEOKA (dti0771@mail4.doshisha.ac.jp)

^{††} Tomoya SUZUKI (tsuzuki@mx.ibaraki.ac.jp)

[‡] Seiichi YAMAMOTO (seyamamo@mail.doshisha.ac.jp)

Graduated School of Engineering, Doshisha University ([†])

College of Engineering, Ibaraki University (^{††})

Faculty of Science and Engineering, Doshisha University ([‡])

1-3 Miyakodani, Tatara, Kyotanabe, Kyoto 610-0321,

Japan

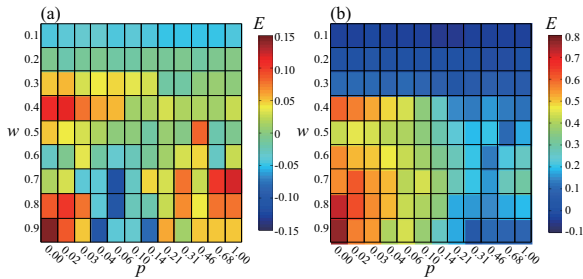


図 1: 提案手法と従来手法間の最良推定精度 $E_m(\theta)$ の差 . (a) 提案手法対大津の閾値法 . (b) 提案手法対並べ替え検定手法 .

係数であり, 要素間の相互作用の強さを表す. また $f(\cdot)$ は本研究では広く研究されているロジスティック写像 (式 (5)) を用いた .

$$f(x_t) = 1 - 2x_t^2 \quad (5)$$

この CML を構成する各要素から得られた時系列データに対し, 相関係数と偏相関係数を用いて構造推定を行い推定値 $\tilde{W}_{i,j}$ を求める. この $\tilde{W}_{i,j}$ に対して提案手法, そして従来より用いられている閾値推定手法である大津の閾値法 [9], 並べ替え検定 [10][11] によりそれぞれ閾値 θ を求め, 結合の有無を判別する. さらに, 式 (6) によって推定精度 $E(\theta)$ を求め, 各手法の比較を行う .

$$E(\theta) = \frac{S \cap \tilde{S}(\theta)}{S \cup \tilde{S}(\theta)} \quad (6)$$

S は正解の要素間結合を表す集合であり推定精度が良いほど, $E(\theta)$ の値は大きくなる. $\tilde{S}(\theta)$ は各手法によって推定された要素間結合の集合である .

次に, 相関係数と偏相関係数の推定精度のうちより高い $E(\theta)$ を最良推定精度 $E_m(\theta)$ として各手法の比較に用いた. 図 1(a) に提案手法の $E_m(\theta)$ と大津の閾値法の $E_m(\theta)$ の差を, 図 1(b) に提案手法の $E_m(\theta)$ と並べ替え検定の $E_m(\theta)$ の差を示す. パラメータ p と w の全領域において常に提案法が優れているわけではないが, 有利な領域が存在することが分かる .

次に, 図 1 の差が統計的に結果に有意であるかを調べるため, 提案手法と従来手法について対応のある片側 t 検定を行った. その際, 帰無仮説 $H_0(1)$ として提案手法は従来手法より有意に優れていないとした. つまり $H_0(1)$ を棄却できれば, 提案手法は従来手法より優れていると支持できる. さらに帰無仮説 $H_0(2)$ では反対に, 従来手法は提案手法より有意に優れていないとした. $H_0(2)$ が棄却された場合は, 従来手法が提案手法に勝ることになる .

この検定結果を図 2 に示す. これらの結果から提案手法は従来手法と比べて良い結果を得られる場合が多く存在することがわかる. 図 2(a) の大津の閾値法に対しては, w が大きく, レギュラーネットワークおよびランダムネットワークに近い状況で提案手法が良い結果となる. 図 2(b) の並べ替え検定法に対しては, ほとん

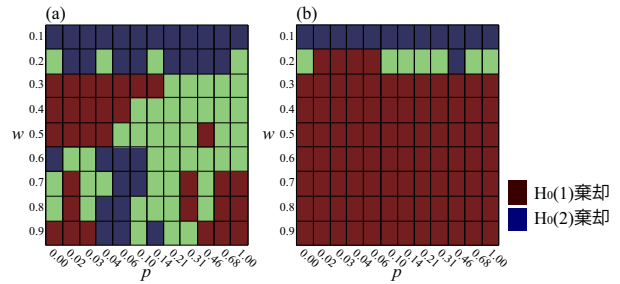


図 2: 提案手法の精度 $E_m(\theta)$ と従来手法 $E_m(\theta)$ 間の t 検定 . (a) 提案手法対大津の閾値法 . (b) 提案手法対並べ替え検定手法 . $H_0(1)$ 棄却: 提案手法は比較手法より良い, $H_1(2)$ 棄却: 比較手法は提案手法より良いことを示す .

ぼ完全に提案手法が有利となる. 逆にたとえ提案法が不利となる領域であっても両手法間の推定精度の差は非常に小さい .

4 まとめ

複雑システムの構造推定を行う際には相互作用の有無を判定する閾値が必要となるが, 本研究では, グレンジャーの因果性に基づいてシステム全体の予測精度を最大化するように, この閾値を定めた. さらに, 従来の閾値決定法である大津法や並べ替え検定法による推定の結果と比較し, 本手法の有用性を確かめた .

なお本研究の一部は, 日本学術振興会科学補助金若手研究 (B)(No.22700227) の援助により行われました .

参考文献

- [1] T. Heimoa, J. Saramaäki, J. P. Onnela, and K. Kaskia: *Physica A* **383**, 147 (2007).
- [2] T. S. Gardner, D. di Bernardo, D. Lorenz, and J. J. Collins: *Science* **301**, 102 (2003).
- [3] S. Frenzel and B. Pompe: *Phys. Rev. Lett.* **99**, 2041101 (2007)
- [4] M. G. Rosenblum and A. S. Pikovsky: *Phys. Rev. E* **64**, 045202 (2001).
- [5] J. Granger: *Econometrica* **37**, 424 (1969).
- [6] Y. Ueoka, T. Suzuki, and S. Yamamoto: *IJMPC* **21**, 1065 (2010).
- [7] D. Watts and S. Strogatz: *Nature* **393**, 440 (1998).
- [8] K. Kaneko: *Formation, Dynamics, and Statistics of Patterns* (World. Sci., 1990).
- [9] N. Otsu: *IEEE Trans. Sys., Man, and Cybernetics SMC-9*, 1, 62 (1979).
- [10] Z. Zhou, M. Ding, Y. Chen, P. Wright, Z. Lu, and Y. Liu: *Brain Research* **1289**, 22 (2009).
- [11] B. R. Borate, E. J. Chesler, M. A. Langston, A. M. Saxton, and B. H. Voy: *BMC Research Notes* **2**, 240 (2009).
- [12] E. N. Lorenz: *J. Atmospheric Science* **26**, 636 (1969).