

部分的な因果構造を利用した複雑システムの大域的因果構造

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

1 はじめに

現実世界には経済システムや脳内の神経回路網などといった複雑システムが多数存在し、内在する各要素は互いに複雑に影響し合うため、全体の振舞いは複雑になり、その結果、予測や制御が困難になる。しかし、システムの振舞いとして観測できる時系列データを用いて各要素間の因果構造を同定できれば、システム全体のモデル化に役立ち、予測や制御を行う上で有利になる。このような動機から、各要素間の振舞いに対して相関係数や偏相関係数などを見積り、因果の有無を判別する方法がある [1]。しかし、これらの推定値が大きいかほど因果関係を有すると判別できるが、その推定結果は因果の有無を判別する閾値に大きく依存するといった問題がある。そこで本研究では、この閾値を最適化する手法を検討する。

なお実際の解析においては、対象とするシステムの因果構造は全く検討がつかないとは限らない。解析対象のシステムを鑑みれば、一部の因果構造を想像できる場合がある。例えば経済システムにおいては、関連企業であれば互いに直接的な影響を及ぼし合っている可能性は高く、逆に全く関連しない企業についても、ある程度は想像できるであろう。さらには、実際に一部の因果構造が明らかなシステムを解析対象とする場合もあるであろう。このような解析対象において、一部の既知の因果構造を手がかりにして、上述の閾値を最適化する。その手法として、一部の既知の因果構造に対して相関係数などの統計的指標を適用し、その推定精度が最良となるように閾値を定める。次に最適化された閾値を用いて、残りの未知の因果構造を推定する。この手法の有用性を検証するために、数値シミュレーションを通じて、一般的なクラス分類手法を用いて閾値を定めた場合との推定精度を比較する。

2 因果構造の推定法

本研究において多変量システムの振舞いを x_i^t とし、 t を時間、 i を要素番号 ($i = 1, 2, \dots, N$) とする。もし要素 a と要素 b の間に因果関係が存在すれば、時系列データ x_a^t と x_b^t は相互作用しながら変動する。これらの時系列データから各要素間の因果構造を推定するには、2つの時系列データ間の相関係数などを算出し、閾値 θ より大きければ要素間に因果関係が存在すると判定できる [1]。しかしその推定精度は、閾値 θ の設定に依存する。

もし一部の要素の因果構造が既知であれば、この閾値 θ を定める指針となる。まず、因果構造が既知の要素間の時系列データに対して、相関係数などの統計量を算出する。詳細は次章で述べる。ここで因果構造が既知であるので、次式により閾値 θ に応じた推定精度 $E(\theta)$ を求めることがで

きる。

$$E(\theta) = \frac{|P \cap N(\theta)|}{|P \cup N(\theta)|} \quad (1)$$

P は真の因果関係 (要素ペアの番号) の集合であり、 $N(\theta)$ は推定された因果関係の集合である。つまり $P = N(\theta)$ のとき、推定精度は最大となる。この推定精度 $E(\theta)$ を最大化するように、最適な閾値 θ_{opt} を決定する。以上の閾値最適化法を、本研究では“提案法”と呼ぶ。以後、未知の要素間の因果構造を判別する際には、 θ_{opt} を用いる。

さらに、一般的なクラス分類手法を応用して閾値 θ を定めることもできる。大津の閾値法 [2] においては、各要素間の統計量を二分するため、クラス内分散を最小化することにより閾値 θ を定める。また、 k 近傍法を応用することもできる。未知の要素間の統計量を算出し、その因果の有無を判別する場合、その統計量に近い要素間の関係を既知の因果構造より探索する。探索された要素間の因果関係の有無に応じて、未知の要素間の因果関係を判別する。

以上の手法の有用性を比較するために、次章において評価実験を行う。

3 評価実験

因果構造を判別するための多変量システムとして、複雑系を模擬する Coupled Map Lattice (CML) [3] を用いる。しかし通常の CML の因果構造は、レギュラーグラフで表現されているため、本研究では、WS モデル [4] を CML に融合し、因果構造をスモールワールドネットワークやランダムグラフに変えられるように改良した (式 (2))。本モデルでは、WS モデルにて結合している要素間に因果関係が存在すると考える。さらに、WS モデルの張り替え確率 p を可変することにより、様々な因果構造を表現できる。

$$x_{i+1}^i = f \left((1 - \epsilon)x_i^i + \frac{\epsilon}{N_i} \sum_{j \in G_i(p)} x_j^i \right) \quad (2)$$

$G_i(p)$ 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 はその集合の要素数である。 x_i^i は i 番目の要素の時系列データである。 ϵ は結合係数であり、要素間の相互作用の強さを表す。また f はダイナミクスであり、本研究では広く研究されているロジスティック写像 (式 (3)) を用いた。

$$f(x_i) = 1 - 2x_i^2 \quad (3)$$

この CML を構成する各要素から得られた時系列データ x_i^t に対し、相関係数、偏相関係数、相互情報量、移動エントロピー [5] を用いて、要素間の因果関係の有無を判別する。また、本研究では要素数 N を 100 とし、全体の 50 [%] の因果構造が既知であるという状況下でシミュレーションを行った。さらに、前章で示した手法によって残りの未知の因果構造を推定し、式 (1) のように推定精度 E を算出した。

図 1~4 において実験結果を示す。各図ともに、提案法による推定精度 E に対する k 近傍法の推定精度 E の比の対数 (左図) および大津の閾値法による推定精度 E との比の対数 (右図) を示している。つまり、0 以上であれば提案法の方が有用である。システムの特性を決めるパラメータ p および ϵ が変化しても、概ね提案法の有用性を確認できる。

Estimating Structure of Complex System Using Partial Known Structure

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

^{††} Tomoya Suzuki (tsuzuki@mail.doshisha.ac.jp)

[†] Seiichi Yamamoto (seyamamoto@mail.doshisha.ac.jp)

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

Department of Information System Design, Faculty of Science and Engineering, Doshisha University (^{††})

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

Japan

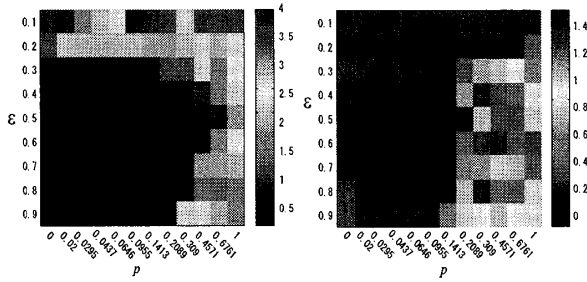


図 1: 提案法による因果構造の推定精度 E の改善量。ただし、要素間の統計量として相関係数を用いた場合。左図は、 $\log(\text{提案法による推定精度 } E / k \text{ 近傍法による推定精度 } E)$ を示し、右図は、 $\log(\text{提案法による推定精度 } E / \text{大津の閾値法による推定精度 } E)$ を示す。

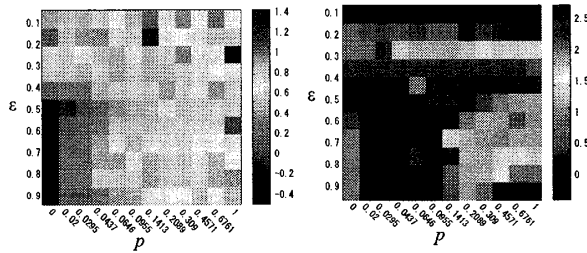


図 2: 図 1 と同様。ただし、要素間の統計量として偏相関係数を用いた場合。

図 1, 2 は、統計量として 2 時系列データ間の相関係数および偏相関係数を用いた場合の結果である。相関係数と偏相関係数は 2 要素間の線形関係を測ることができる。しかし相関係数は、共通の第 3 要素からの影響により疑似相関が発生する弱点がある。いずれの統計量を用いても、 p が大きい場合に提案法による改善量が大きい。

さらに図 3, 4 において、相互情報量および移動エントロピー [5] を用いた場合の結果を示す。これらの統計量は 2 要素間の非線形関係も測ることができる。しかし相互情報量は、相関係数と同様に疑似相関の問題を持つ。相互情報量は式 (4)、移動エントロピーは式 (5) で定義される。

$$M_{i,j} = \sum_t p(x_t^i, x_t^j) \log \frac{p(x_t^i, x_t^j)}{p(x_t^i)p(x_t^j)} \quad (4)$$

$$T_{i \rightarrow j} = \sum_t p(x_{t+1}^j, x_t^i, x_t^j) \log \frac{p(x_{t+1}^j | x_t^i, x_t^j)}{p(x_{t+1}^j | x_t^j)} \quad (5)$$

ここで、 x_t^i は要素 i の時間 t の値であり、 $p(x_t^i, x_t^j)$ は同時確率、 $p(x_{t+1}^j | x_t^i, x_t^j)$ は条件付き確率である。同時確率の式として、式 (6) を用いる。

$$p(x_t^i, x_t^j) = \frac{1}{L} \sum_{\nu} \Theta \left(F_s \gamma - \left| \begin{pmatrix} x_t^i \\ x_t^j \end{pmatrix} - \begin{pmatrix} x_{t-\nu}^i \\ x_{t-\nu}^j \end{pmatrix} \right| \right) \quad (6)$$

Θ はステップ関数であり、 L は時系列データ長、 F_s は時系列データのフルスケールである。また分解率として γ を設定する必要がある。閾値 θ と同様に、因果構造の推定精度はこの γ にも依存する。本研究では、 θ に関する提案法と同様に、既知の因果構造の推定精度を最大化することにより γ の最適化も行った。相互情報量の結果を図 3 に、移動

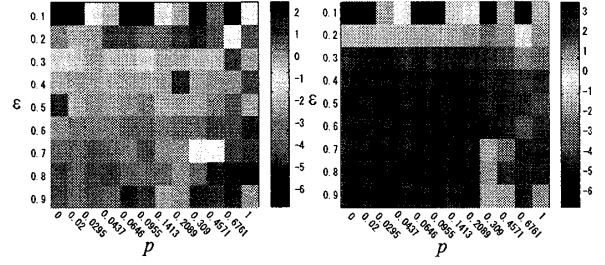


図 3: 図 1 と同様。ただし、要素間の統計量として相互情報量を用いた場合。

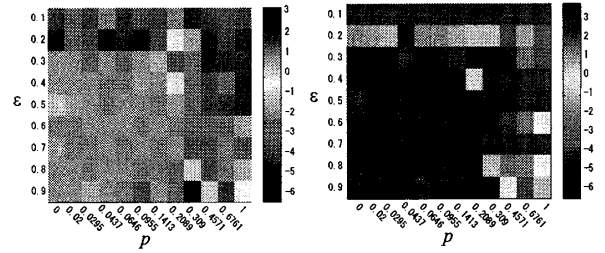


図 4: 図 1 と同様。ただし、要素間の統計量として移動エントロピーを用いた場合。

エントロピーの結果を図 4 に示す。これらの場合においても、提案法による推定精度の向上を確認できる。特に大津の閾値法と比べて、提案法による推定精度は高い。

4 まとめ

本研究では、複雑システムの因果構造を推定する問題において、因果構造の有無を区別する閾値の最適化手法について検討した。提案法として、既知の因果構造の推定精度を最大化することにより閾値 θ や非線形統計量の γ の最適化を行った。さらに、これらのパラメータを用いて未知の因果構造を推定し、その推定精度について k 近傍法や大津の閾値法などのクラス分類手法との比較を行った。その結果、複雑システムの特性によらず、提案法の有用性を確認できた。

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.20700217) の援助により行われました。

参考文献

- [1] Y. Ueoka, T. Suzuki, T. Ikeguchi, Y. Horio: "Efficiency of Statistical Measures to Estimate Network Structure of Chaos Coupled Systems," Proc. of NOLTA'08, 2008.
- [2] N. Otsu: "A Threshold Selection Method from Gray-Level Histograms," IEEE Trans. Sys., Man, and Cybernetics, SMC-9, No.1, pp.62-66, 1979.
- [3] 金子邦彦, 津田一郎, "複雑系のカオス的シナリオ," 朝倉書店, 1996.
- [4] D. Watts and S. Strogatz: "Collective dynamics of 'small-world' networks." Nature, vol.393, pp.440-442, 1998.
- [5] T. Shreiber: "Measuring Information Transfer," Physical Review Letters, vol.85, pp.461-464, 2000.