

細胞などの要素間相互作用の関係性をデータ駆動的に解明

～データ科学的に困難であった要素間の主従関係を同定する有力な手法として期待～

ポイント

- ・細胞などの要素間における多体の相互作用をデータ科学的に解明する手法を開発。
- ・データから因果関係を推論する従来の理論を一般化し、主従関係をより正しく評価することに成功。
- ・複雑な要素間における多体の相互作用を分析する有力なデータ駆動型手法として期待。

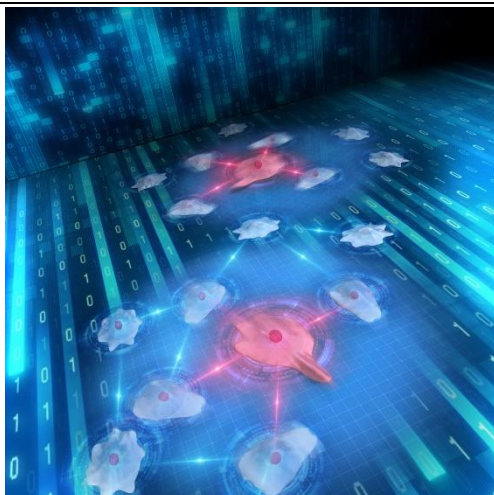
概要

北海道大学電子科学研究所の小松崎民樹教授の研究グループは、カリフォルニア大学デービス校のジェームズ クラッチフィールド教授らと共同で（人間，鳥，魚，細胞などの）要素間の多体の相互作用を、因果関係を推定する情報理論^{*1}を改良することで、従来法に比して主従関係をより正しく評価できるとともに、2つの要素の軌跡データだけを用いて分析できることを見出しました。

ある細胞と別の細胞の主従関係を推定する場合、それら2つの細胞の軌跡データなどを用いて評価されてきました。つまり、主従関係を理解するには、すべての細胞対の軌跡データを調べることとなります。しかしながら、2つの細胞の振る舞いを決める因子が、その2つの細胞以外にも、第3の細胞が介在する状況なども考えられ、主従関係や因果関係における“原因”と“結果”を解析するためには、単純な対の組み合わせで表現できない多体の相互作用から成り立っています。そのため、多体のあいだの因果関係を推定することは要素間の組み合わせの数が膨大になり、データ科学における難問でした。そこで、研究グループは、情報理論において因果関係を推定する移動エントロピーと呼ばれる量を細分化した新しい情報量に着眼し、生物等の集団運動を模倣する数理モデルに基づいてこの問題を考察しました。さらに、移動エントロピーに含まれる相乗情報量の振る舞いを解析することで、要素間の多体の相互作用が推定できる可能性を示しました。

本研究成果により、従来の移動エントロピーよりも、背後の因果関係をより正しく評価が可能となり、系の振る舞いを評価する式自身が非自明な複雑な要素間の多体の相互作用を分析する有力なデータ駆動型手法として期待されています。

なお、本研究成果は2022年2月9日（水）公開のScience Advances誌に掲載されました。



リーダー細胞（赤）を取り巻くフォロワー細胞（白）の多体の相互作用の概念図。

手前のモデルは、隣接する近傍のフォロワー細胞は互いに影響を及ぼしあっているのに対し、奥のモデルでは、フォロワー細胞はお互いの影響が無視できるほど小さい。このような多体の相互作用の違いを二対の細胞の運動の軌跡データのみから識別する問題。

【背景】

集団行動は、人間だけでなく、鳥、魚、細胞を含む様々な生物階層の集団において存在します。多くの場合、人間、鳥、魚、細胞などの各要素は相互に作用しあい行動を取っています。集団行動における相互作用の関係を理解することは、現代社会においても重要といえます。例えば、どのように車の動きが他の車の動きに影響を受け、または与えているのかを理解することで、交通の流れをより効率よく制御したり、魚がどのように群れの中で相互に作用しあっているかを理解したりすることで、魚の群れを制御して、狙った魚をほかの魚を傷つけることなく効率的に釣ることができるかもしれません。

ある細胞と別の細胞の主従関係を推定する場合、それら2つの細胞の軌跡データなどを用いて評価されてきました。主従関係の“見取り図”を作成する際には、すべての細胞対の軌跡データを調べることに なります。しかしながら、2つの細胞の振る舞いを決める因子が、その2つの細胞以外にも、第3の細胞が介在する状況なども考えられ、因果関係における“原因”と“結果”を解析するためには、単純な二対の組み合わせで表現できない多体の相互作用から成り立っています。しかしながら、多体のあいだの因果関係を推定することは要素間の組み合わせの数が膨大になるため、データ科学における難問でした。

集団運動において、(細胞、鳥、人などの)要素間の主従関係は二対の要素の軌跡データからしばしば推定されてきました。例えば、 n 個の要素から成る系を考えてみましょう。もし、その系の主従関係が単純な二対の要素の軌跡データで推定できるのであれば、 ${}_nC_2$ 通りの要素の二対の軌跡を解析することになります(例えば、 n が100の場合、 ${}_{100}C_2 = 4950$ 通り)。しかしながら、仮に、二対の振る舞いが、それ以外の第三、第四、...の要素の影響も受けているとすると、二対の単純な組み合わせを調べるだけでは主従関係を推定することは困難となり、考え得るすべての組み合わせの要素の情報を調べなければならないこととなります。例えば、10個の要素が関与しているとする、 $n=100$ 、 $m=10$ 、 ${}_{100}C_{10} = 17,310,309,456,440$ 通りの場合の数を調べることに、多体の相互作用をデータ科学的に有限のデータ量から評価することは極めて困難な問いとなります。

データから因果関係を推定する手法が情報理論において移動エントロピーと呼ばれる量が提案され、これまでも様々な分野において広く用いられてきました。これは2つの変数の時間変化を入力として、一方の変数のそれまでの履歴を知ること、もう一つの変数の次の時刻での出力がどれくらい予測し得るかを定量することで情報の流れ(=因果関係)を推定するものです。

共同研究者であるクラッチフィールド教授らは、近年、移動エントロピーには欠陥があること、送受信者間で安全な通信を実現するための鍵(パスワード)を共有する情報理論的暗号の考えを導入し、移動エントロピーを本来、定量したかった固有の情報の流れのほかに、相乗情報量と呼ぶべき量に近似的に分解できる可能性を示唆しましたが、どのような新たな知見を与え得るかについては不明瞭なままでした。

本研究では、生物等の集団運動を模倣するモデルを構成し、移動エントロピーを分解することで、背後の多体の相互作用を推定できるかを調べました。

【研究手法】

細胞などの集団運動を模倣する主従関係モデル(図1)を構成し、リーダー細胞、リーダーに追随するフォロワー細胞を導入し、フォロワー細胞の動きがリーダー細胞だけでなく近傍のフォロワー細胞からも影響を受ける多体の相互作用が無視できない状況(図2)を構成し、移動エントロピーに内在している固有の情報の流れと相乗情報量を解析しました。

【研究成果】

リーダー細胞からフォロワー細胞への移動エントロピーに含まれる相乗情報量を調べた結果、フォロワー細胞同士は互いに相互作用しないモデルではフォロワー細胞の数に依存しないこと（図3左）、フォロワー細胞が互いに相互作用するモデルではフォロワー細胞の数が増えるにつれて、相乗情報量は小さくなること（図3右）がわかりました。この相乗情報量は、リーダー細胞とフォロワー細胞の両方の履歴情報がフォロワー細胞の次の時刻の運動の振る舞いをどれくらい予測できるかを定量しています。すなわち、後者のモデルでは、フォロワー細胞はリーダー細胞に加えて他のフォロワー細胞の影響も受けるため、フォロワー細胞の数が増えるにつれて、リーダー細胞からフォロワー細胞の動きを予測する性能が急激に減少します。このことは、リーダー細胞とフォロワー細胞のあいだの2体の軌跡データだけから構成される相乗情報量を評価することで、第3のフォロワー細胞の影響を推定できることに対応しています。

このほか、相乗情報量と相互作用ネットワークにおける次数（相互作用の入出力）の関係なども明らかにし、多体の相互作用を二対の情報量から推定できる可能性が高いことを明らかにしました。

【今後への期待】

本研究成果により、それまで組み合わせの数が爆発することで困難であった多体の相互作用存在下における主従関係を二対の要素の軌跡データから推定できることが示唆されました。また、提案された理論的枠組みと用いられたデータ分析法は、二対の要素の軌跡データの振る舞いに影響を与えるものが、第三の要素（例：その二対の細胞以外の別の細胞）である必要は必ずしもなく、要素間の多体の相互作用を分析する有力なデータ駆動型手法として期待されます。

【謝辞】

本研究は新学術領域シンギュラリティ生物学、JST/CREST「情報計測」、北海道大学創成研究機構化学反応創成研究拠点 WPI-ICReDD MANABIYA、物質・デバイス領域共同研究拠点等の支援を受けて実施されました。

論文情報

論文名	Modes of information flow in collective cohesion (協同的凝集現象における情報流れのモード)
著者名	Sulimon Sattari ¹ , Udoy S. Basak ^{1,2} , Ryan G. James ^{3,4} , Louis W. Perrin ⁵ , James P. Crutchfield ⁴ , and Tamiki Komatsuzaki ¹ (¹ 北海道大学電子科学研究所, ² パプナ科学技術大学・バングラデシュ, ³ Reddit, Inc・米国, ⁴ カリフォルニア大学デービス校・物理学科, 複雑系科学センター・米国, ⁵ レンヌ高等師範学校・フランス, ⁶ 北海道大学創成研究機構化学反応創成研究拠点, ⁷ 北海道大学総合化学院)
雑誌名	Science Advances
DOI	10.1126/sciadv.abj1720
公表日	2022年2月9日(水)(オンライン公開)

お問い合わせ先

北海道大学電子科学研究所・創成研究機構化学反応創成研究拠点 教授 小松崎民樹（こまつぎきたみき）

T E L 011-706-9434 F A X 011-706-9434 メール tamiki@es.hokudai.ac.jp

U R L <https://mlns.es.hokudai.ac.jp/>

配信元

北海道大学総務企画部広報課（〒060-0808 札幌市北区北8条西5丁目）

T E L 011-706-2610 F A X 011-706-2092 メール jp-press@general.hokudai.ac.jp

【参考図】



図 1. 1つのリーダー細胞 (L), 2つのフォロワー細胞 (F) の相互作用モデルのネットワーク表現。

矢印は現時点での運動が次の時刻の運動に影響を与える相互作用を表現し、太さはその大きさを表現している。解析にはリーダー細胞がフォロワー細胞へ与える影響の大きさはその逆に比して、1%大きく設定している。

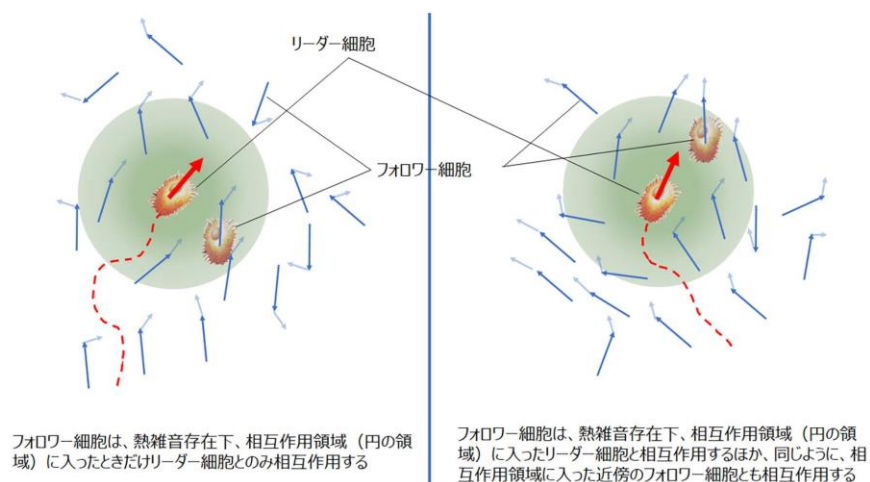


図 2. リーダー細胞とフォロワー細胞の相互作用。

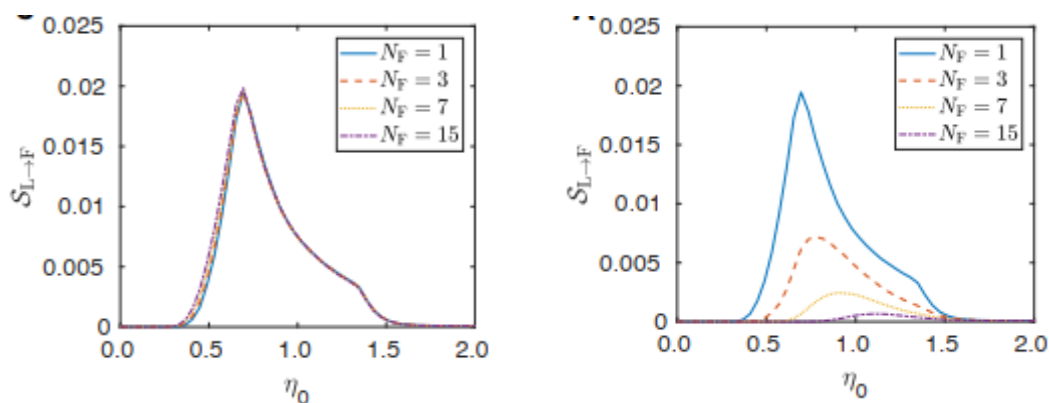


図 3. リーダー細胞からフォロワー細胞への移動エントロピーに内包される相乗情報量を熱揺らぎの大きさ η_0 に対してプロットしたもの。各線はリーダー細胞 1 つに対するフォロワー細胞数の違いを表している。