

CASE 1

深層学習の特徴抽出でがんの新たな知識を発見

機械学習の一つである深層学習の能力を遺伝子解析の分野でも存分に活用したい。

だが、ゲノムなどの遺伝子のデータは高次元変数データであり、

深層学習が得意とする画像などのパターンではない。

そこで遺伝子のデータをうまく変換して画像のように見せ深層学習にかける離れ技を練り出し、

たとえばがんの遺伝子のデータを入力させてがんの種類を判別できるようになった。

さて、では、深層学習はいったい何を見て判別しているのだろうか？それをどうすれば分析できるのだろうか？

そして実際にがんを分析したとき、いったい何が発見できるのだろうか？

私たちは、がん患者からの分子データでがんの発生メカニズムを解明し、患者ごとにがんの個性を見分け最適な治療法を選ぶ方法を研究している。ここでいう分子データとは、ゲノムやそこから転写されるRNAの総体であるトランスクリプトーム、細胞内で発現している全タンパク質を指すプロテオーム、メタボローム（生体内に含まれる代謝物質の総体）、遺伝情報を修飾するエピゲノムなどで、これらを網羅的にまとめて研究する学問をオミクスと言う。がんの個性はこの多くのオミクス間の複雑な関係で決まるので、従来の統計学では扱いにくい。

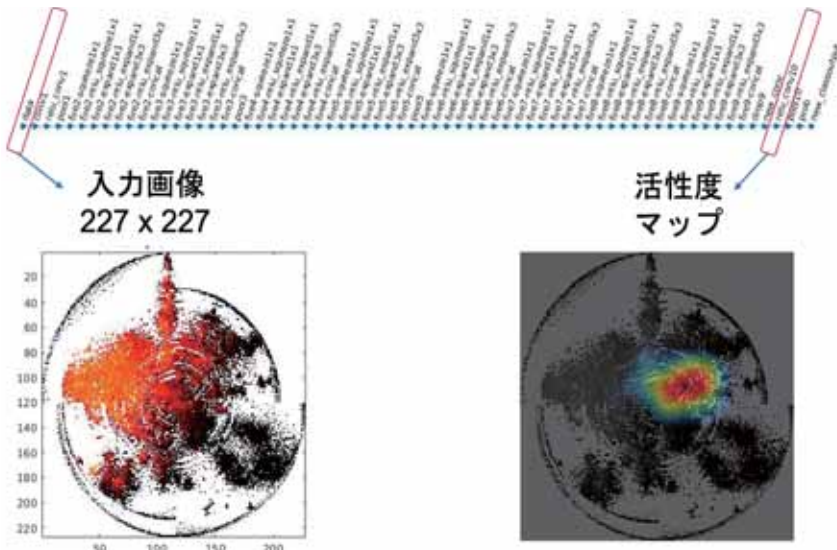
深層学習は、非線形関数の中間層を複数持つ深層ニューラルネットワーク（DNN）に基づき、複雑なパターンを扱え、画像処理などに優れていることから着目していたが、オミクスは画像ではなく、高次元のデータであることから良い方法がないかと試行錯誤した結果、私はオミクスデータを2次元で表現できれば良いはずだとひらめき、多くの検体間で似た挙動をする分子どうしが近くになるように分子をマップし直せばよいと気づいた。こうして画像にしたオミクスデータを深層学習で扱い、がん種を高精度に判定できた（2019年発表 DeepInsight法）。ではDNNは何を見てがんを見分けているのだろうか？

今回、私たちはそれを発見する DeepFeature法を考案した。DNNでは、学習後、出力層近く

の中間層（図右側）にさまざまな特徴が集約されており、それらを総合して画像を判別する。この層を分析することが鍵であると考えた。分析にはCAM（Class Activation Mapping）という技術を用いた。学習後に画像を入力すると（図左端）、この中間層上の (x,y) 座標での出力 $f_k(x,y)$ が特徴 k ごとに計算される（図右側の「活性度マップ」）。さらに、各特徴 k が判別対象（例えば肺がん）を支持する度合いを表す重み w_k もDNN内で計算されている。そこで特徴ごとに $f_k(x,y) \cdot w$ を計算し、最後に全特徴で足し合わせる。すると、肺がんと判定されたとき、入力画像の各画素が寄与する度合い、つまりどの画素を見て判定したかがわかる。こうして構築した DeepFeature法を10種のがんを判別予測する実験に適用すると、上皮間葉転換や凝固、血管新生、低酸素、炎症反応など、がんによく関わる遺伝子を見ていることがわかった。さらに、コラーゲンに代表される細胞外マトリックス構造やタンパク質のチロシン残基のリン酸化を介した受容体チロシンキナーゼシグナル伝達、Gタンパク質共役型受容体（GPCR）リガンド結合などのシグナル経路を見ていることも新たに発見できた。

深層学習には結論に至る道筋がわからないという「ブラックボックス」問題があるが、本研究を糸口にそれを脱却し、経緯を発見することが夢でなくなるだろう。今後、複雑でダイナミックながん化のメカニズムを解明し、患者ごとにがんの細かい「顔」の違いをみて治療を決めることを目指している。

本研究成果は A. Sharma *et al.*, *Briefings in Bioinformatics* 22, bbab297 (2021) に掲載された。（2021年8月19日プレスリリース）



図：がん患者のがん組織からの非画像のオミクスデータを画像のように変換し、深層学習の入力画像とする（左端）。この画像を入力すると、深層ニューラルネットワーク（上側）の左から右へ中間層上のニューロンが発火していく。出力層に近い層が特徴を表現する層（右端側）で、その活性度の様子（「活性度マップ」、色の着いた部分）から、どのような特徴を見ているかがわかる。論文から改変。

CASE 2

宇宙の有機物の塵を作る

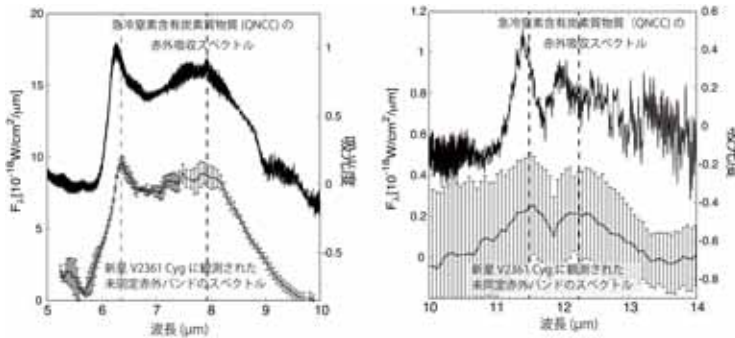
赤外線天文衛星は、大気に妨げられず、

さまざまな天体や銀河の星間物質の赤外スペクトルを取得することができる。

その観測によって、3-20 μm の波長域に複数の幅広いバンド放射から成る「未同定赤外バンド」が、天体や銀河にあまねく存在することが明らかになった。

1973年に初めて発見された未同定赤外バンドは、その担い手が有機物に起因することが知られているが、発見から約半世紀が経とうとする今も、その有機物がどのような物質であるか同定するための試みが続けられている。

本記事では、新星に観測される未同定赤外バンドの性質を再現する有機物の塵の合成実験の最新結果を紹介する。



老いた星は、重元素を含む恒星風を星間空間に放出し、「有機物の塵」の生成現場となる。新星もその一つで、新星が放出するガス中には、太陽風の数百倍以上高い濃度の窒素が含まれることがある。星間物質全体の中で、新星に起源を持つ物質の割合は多くない。しかしながら、未同定赤外バンドの放射を示す新星が幾つも観測されており、出来立ての有機物の塵の素性を観測する上で、新星は貴重な天体現象である。

そこで、地上で星間物質の合成を模擬し、観測される未同定赤外バンドの特徴を再現しようとする試みが行われている。その歴史は長く、本研究で使用したマイクロ波を用いたプラズマ発生装置は、1970年代後半に、電気通信大学の坂田朗助手（当時）が製作したものである。当時、この実験装置を用いてメタンガスをプラズマ状態から急冷凝縮して得られる「急冷炭素質物質（Quenched Carbonaceous Composite; QCC）」は、星間塵の光学的な性質を議論する国際的な場で重要な役割を果たした。2008年に電気通信大学の和田節子准教授が退職される際、本学天文教室にこの装置を移管した。

急冷炭素質物質が炭素と水素から成る有機物の塵であるのに対し、われわれが今回取り組んだのは、窒素を含む有機物の塵の合成である。炭化水素の固体試料と希薄な窒素ガスをマイクロ波加熱し得られたプラズマを急冷凝縮することで得られる有機物の塵の赤外線吸収スペクトルが、新星に観測される未同定赤外バンドの特徴をよく再

現することを発見した。われわれは、その有機物の塵を「急冷窒素含有炭素質物質（Quenched Nitrogen-included Carbonaceous Composite; QNCC）」と名付けた。X線吸収端近傍構造分析の結果、急冷窒素含有炭素質物質には窒素がアミンの形態で含まれていることがわかった。

急冷窒素含有炭素質物質は、炭素質コンドライト隕石中から抽出される不溶性有機物とも類似した赤外特性を示す。太陽系の有機物の起源として、年離れた星が作る有機物の塵が過酷な星間空間を長期に旅し生き残り新たな太陽系の誕生の現場に取り込まれるという道筋もあり得るのではないかと。2015年以降、国際宇宙ステーションきぼう実験棟簡易曝露実験装置 ExHAM（エクスハム）を用いて、実験室で合成した急冷窒素含有炭素質物質などを太陽近傍の宇宙環境に曝露し、有機物の塵が獲得する変性過程を調べる実験に着手した。回収試料と、隕石中や実際に宇宙から採取した物質に含まれる太陽系の始原的な有機物を比較し、先の仮説の検証に挑みたい。

天文学的手法と地球科学的手法を総動員させて、究極的には、太陽系に生きる我々の存在が広く長い宇宙の歴史の中ではありふれたものであるかどうかを知る、そうした目標に急冷窒素含有炭素質物質が役立つことを願う。

本研究成果は I. Endo *et al.*, *The Astrophysical Journal* 917, 103 (2021) に掲載された。

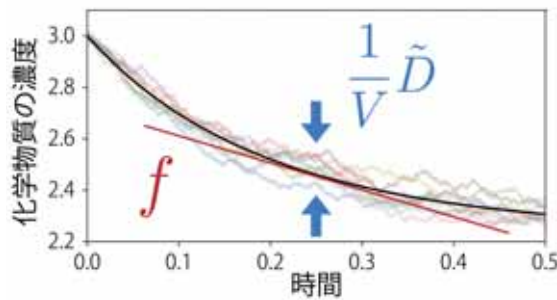
(2021年8月26日プレスリリース)

図：(左) 急冷窒素含有炭素質物質 QNCC の赤外吸光度スペクトルと新星 V2361Cyg に観測される未同定赤外バンドの比較。(右) 急冷窒素含有炭素質物質の合成の様子。

CASE 3

新たな熱力学法則 マクロな化学反応系における

マクロな化学反応系は熱力学・化学熱力学の法則に従う。
たとえば閉じた系での化学反応は、ギブスの自由エネルギーが減る方向に進むことが知られている。
近年、ゆらぎの影響が無視できない小さな系での熱力学の研究が行われ、
ゆらぎの指標と自由エネルギーの減少スピードとの間に
トレードオフ関係が成り立つという新法則が見つかった。
このような新法則はゆらぎが見えないマクロな化学反応系では成り立たないのだろうか？
この疑問に対して成り立つと肯定的に答えたのが本研究である。



スケールされた拡散係数 \tilde{D}
(系のサイズ V に反比例する
ゆらぎの大きさにおける係数)

反応による濃度変化の傾き f

エントロピー生成率 $\dot{\Sigma}$

熱力学的不確定性関係

$$\dot{\Sigma} \tilde{D} \geq f^2$$

マクロな系の熱力学は、適用範囲の広さから現代の科学の中で重要な位置づけをされている。特にマクロな化学反応系の化学熱力学は、化学反応の進む方向を決定するという重要な役割を担っている。

従来のマクロな系の熱力学は平衡状態間の遷移を扱うことが多く、平衡から離れた非平衡状態を扱う部分は未完成とされてきた。そこで近年、ブラウン運動などの確率的なノイズに代表される「ゆらぎ」の影響が無視できない小さな系を用いて、非平衡状態の熱力学を研究する分野が進展している。小さな系では確率的な記述が可能なことから、分散のようなゆらぎの指標が導入可能で、このようなゆらぎの指標を用いた新たな熱力学法則が見つかった。

新たな熱力学法則の中でも最近では熱力学的不確定性関係とよばれる関係式の研究が進んでいる。この関係式は、ゆらぎの指標とエントロピー生成率（閉じた化学反応系においてはギブスの自由エネルギーの減少スピードに相当）の間に、一方が大きければ他方は小さくなるというトレードオフ関係があることを表している。またこの関係式により、状態間の遷移にかかる時間の制限を与えることも知られている。一方で、このような関係式が、ゆらぎが見えないマクロな化学反応系で成り立つかは定かではなかった。

この疑問に対してわれわれは、系のゆらぎがマクロな系で見えなくなる理由に相当する、系のサイズに反比例してゆらぎの大きさが減る関係に着目した。系のゆらぎはマクロな量で記述できない一方で、反比例関係における係数についてはマクロな量だけで記述可能である。われわれはこの係数を内在的なゆらぎの量とみなし「スケールされた拡散係数」とよび、これを用いて熱力学的不確定性関係と同様の関係式を、マクロな系の化学熱力学において示すことに成功した。それによると、閉じたマクロな化学反応系において、ギブスの自由エネルギーの減少スピードと「スケールされた拡散係数」の間にトレードオフ関係があることが新たにわかった。

本結果は従来の熱力学法則よりも詳細に化学反応の速度への制限を加えるものであり、さまざまな応用が考えられる。たとえば反応速度が重要な生化学反応において、生体内の情報処理の正確性と熱力学的なコストの間にトレードオフ関係があるとされてきたが、この関係に対する普遍的な理解のために本結果は新たな視点を与えてくれるだろう。

本研究成果は、K. Yoshimura and S. Ito, *Phys. Rev. Lett.* **127**, 1606018 (2021) に掲載され、Editors' Suggestion に選ばれた。

(2021年10月12日プレスリリース)

図：化学反応で化学物質の濃度が変化する際、系のサイズ V に反比例するゆらぎが内在している。われわれはこの内在するゆらぎの量（スケールされた拡散係数）とエントロピー生成率の間にトレードオフ関係が成立することを示した。