AIは何を考えているのか

山﨑将太朗

2016年以降，人工知能（AI）という言葉を盛んに耳にするようになった．実際に，顔認証や言語の翻訳などAIに日常的に触れている人も多いと思う．このようにAIが発達した大きなきっかけは，機械学習法の一つである深層学習（ディープラーニングまたは多層ニューラルネットワーク）の登場であり，それを後押ししたのがビッグデータとも呼ばれる非常に大規模なデータの集積である．生命科学の分野でも，長年の情報の蓄積やさまざまな技術の発達によって情報が大規模化してきており，その情報という資源を有効活用するための手法としてAIが着目されることも少なくない．一方で，研究や開発にAIを用いる際の問題の一つが，AIがどういう過程を経て，どういう根拠に基づいて結果を導いたのかが不明瞭である点，すなわち，何を考えていたのかよくわからない点である．

データから現象の予測や分類をすることについては，AIは非常に優秀であり，たとえば，遺伝子マーカーデータからの作物形質の予測や，アミノ酸配列からのタンパク質の立体構造の予測，特定のタンパク質に結合する他のタンパク質や化合物の予測，画像データからの細胞種の判別，ゲノムの塩基配列からのプロモーター領域の予測など，さまざまな試みに用いられている[1,2]．これらのAIは，「このような遺伝子マーカーのパターンの個体だとこういった形質を示す傾向があるようだ」といった法則を，大規模なデータの中から学んでいくことで，複雑な現象の予測や分類を可能にし，時には人を凌駕する精度を見せる．AIが人を凌駕するということは，人が気づいてない法則を学んでいるということであり，AIの考えた過程を我々が知ることができれば，これまではわからなかった複雑な法則や，人が思いつかなかった新たな法則を発見できる可能性がある．しかし，精度の高いAIほど，中身はあまりにも複雑でブラックボックスと化してしまい，何を考えて判断したのかがわからなくなる（解釈性が低い）傾向がある．特に，非常に高い精度を誇り，さまざまな派生法による広い応用力をも持つ深層学習は解釈性が低い構築法の代表格であり，学習後のAIから，特定の形質に重要な遺伝子多型，特定のタンパク質に結合する化合物に共通する構造，ガン細胞の形態上の特徴，プロモーターの機能に重要な配列などを知ることは難しいのが現状である．もちろん，精度の高いAIは，正確な予測による育種のシミュレーションや，阻害剤候補の選抜，顕微鏡写真の自動分類，遺伝子情報の整備など研究や開発の補助や高速化に非常に有用である．しかし，それらの分野では，現象が起こる要因，あるいは予測や分類の背景にあるメカニズムが求められることも多い．「なぜかわかりませんが，このような予測になっています」では困るのだ．正確な予測だけではなく，その要因もわかることによって，説得力が増すだけではなく，要因を対象にしたさらなる研究やAIの妥当性の判断も可能になる．したがって，高い精度を保ちつつ，解釈性を上げることが，AIのさらなる発展に向けての大きな課題の一つとなっている．

精度と解釈性をある程度は両立した既存のAIの構築法として，ランダムフォレストなどの手法がある．しかし，その精度と解釈性は十分に高いとは言えず，より良い手法を開発するため，さまざまな試みが行われている．たとえば，解釈性の高い既存の手法の組合せや，まったく新しい概念などを取り入れることで新しい手法が次々と開発されており，判断基準がはっきりとしたAIも登場してきている[3]．加えて，解釈性が非常に乏しい一方で，非常に高い精度を有する深層学習について，そのブラックボックスの中身を理解するための研究も行われている．たとえば，改変すると予測結果に大きく影響するような重要な特徴を探したり，予測過程を逆にたどることで重視した特徴を探したり，特に注意すべき特徴をAIが学習し提示できる機構を追加したりすることで，AIが何に着目して判断したのかが少しずつわかるようになってきている[4]．実際に，AIの応用が盛んな医療分野の最新の研究では，電子カルテのデータを用いて院内死亡率や再入院の有無，入院の長期化などを高精度で予測し，その予測において重要な判断材料となった注意すべきカルテ上の単語や文章，計測値などの情報を医師にわかりやすく提供するAIが開発されている[5]．現在，世界は第三次AIブームの最中にあり，その技術は飛躍的に進歩している．いずれは，AIがどんな法則を学び，何を考え判断しているのかがわかる未来も訪れるだろう．

1) Ma, W. *et al.*: *bioRxiv*, https://doi.org/10.1101/241414 (2017).
2) Tian, K. *et al.*: *Methods*, **110**, 64 (2016).
3) NEC 異種混合学習：https://jpn.nec.com/ai/analyze/pattern.html (2018/9/21).
4) Montavon, G. *et al.*: *Digital Signal Process.*, **73**, 1 (2018).
5) Rajkomar, A. *et al.*: *Digital Med.*, **1**, 18 (2018).

著者紹介　奈良先端科学技術大学院大学先端科学技術研究科バイオサイエンス領域（博士研究員）　E-mail: s-yamasaki@bs.naist.jp