

生成 AI は化学に何をもたらすか?

化学分野における生成 AI の可能性

今や AI (人工知能) は科学に欠かせない存在となっている。
特に近年は、自然言語で人と対話をしたり、文章や画像などを作成してくれる「生成 AI」の発展と普及が目覚ましい。
今後、生成 AI は化学分野にどのような影響を及ぼすのだろうか。
研究者に話を伺った。

2024 年のノーベル化学賞と物理学賞を
独占した AI

2024 年のノーベル物理学賞には、AI の基礎となる「人工ニューラルネットワーク」による「深層学習 (ディープラーニング)」を創製したトロント大学のジェフリー・ヒントン名誉教授とプリンストン大学のジョン・ホップフィールド名誉教授が選ばれた。また、ノーベル化学賞には、AI を使って未知のタンパク質を設計したワシントン大学のデビッド・ベイカー教授と、タンパク質の複雑な立体構造を高精度で予測したグーグル・ディープマインド社のデミス・ハサビス博士 (最高経営責任者) およびジョン・ジャンパー博士 (上席研究員) の 3 名に贈られた。特にグーグル・ディープマインド社の両名が開発したタンパク質の立体構造を予測するプログラム「AlphaFold2」には、米 OpenAI 社が 2022 年 11 月に公開し、瞬く間に世界中に広まった言語生成 AI「ChatGPT」と同じアルゴリズム「Transformer」が使われていたことから、幅広い層から大きな関心が寄せられた。

Transformer など生成 AI のアルゴリズム

は、空間を 3 次元までしか認識することができない我々とは異なり、多次元空間を扱える点が特徴だ。例えば、ChatGPT の場合、Transformer を使い文章を構成する単語をすべてベクトルに変換し、数千次元を超える多次元のベクトル空間に配置している。その際、関係性が強い単語同士ほど空間上の近い位置に配置されるため、単語同士の関係性を広範囲にわたって瞬時に把握することができる。それにより、ChatGPT は長く複雑な文章であっても言葉と言葉のつながりをより正確に把握し、自然な文章を生成することができるのだ。

「生成 AI」とは?

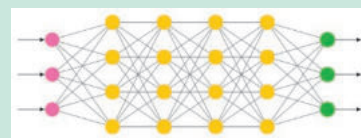
そもそも生成 AI とは、人が与えた大量のデータを学習し、得た情報を基に新たに文章や画像、音声などを生成する AI のことだ。中でも言葉や文章を学習させることで、人が書いたような文章を作成したり、人との自然な対話が可能な生成 AI が、ChatGPT に代表される「言語生成 AI」だ。ChatGPT の登場は人並みの自律思考が可能な汎用 AI の出現に近いことを我々に予見

●人工ニューラルネットワークによる深層学習のしくみ

生成 AI はデータ間の関係を表す「モデル」によって言語や画像を生成する。モデルは、脳神経回路のような多層構造をしており、「人工ニューラルネットワーク」と呼ばれる。人工ニューラルネットワークはネットワークを構成する「ノード」と、ノード間の結びつきの強さを表す「パラメータ」で構成されている。モ

デルは人工ニューラルネットワークの多層構造が深ければ深いほど、つまりパラメータ数が多ければ多いほど、また、読み込ませるデータ量が多ければ多いほど性能が高まる。そのため、現在、多くのモデルの学習に使われているのが、「深層学習 (ディープラーニング)」だ。深層学習では、モデルに大量の「教師データ」

を学習させることでモデルの性能を高めている。モデルにあらかじめ学習させておくことを「事前学習」という。



人工ニューラルネットワーク

させ、世界中に大きな衝撃を与えた。

大学レベルの知識は十分備える 生成 AI

生成 AI は、理論、実験、情報科学に次ぐ、第4の科学手法であるデータ科学に分類される。まず、19世紀後半に、化学結合や分子構造から物性を理解する学問が始まった。そして、20世紀に入り、情報科学の急速な発展により、第一原理計算や分子動力学によるコンピュータ・シミュレーションが大きな役割を果たすようになった。そして、情報科学を基に発展してきたのが、データ科学だ。データ科学の手法としては、「ケモインフォマティクス」や「マテリアルズ・インフォマティクス (MI)」がある。ケモインフォマティクスは化学分野に特化したデータ分析手法、MI は物質・材料科学に特化したデータ分析手法だ。ケモインフォマティクスや MI には、すでに知られている物質・材料の物性データを学習させることで、未知の物質・材料の物性を予測する順問題と、欲しい物性を基に、物質・材料の化学組成や結晶構造を提案させる逆問題の2つのアプローチがある。いずれも大量のデータをコンピュータに学習させることで性能向上を図っている。一方、生成 AI は幅広い分野に適用可能な汎用性を持ち、自然言語による人との円滑なコミュニケーションが可能な点が特徴だ。

「最近の生成 AI は大学院の講義レベルの化学であれば、十分な知識を備えているように見受けられます。簡単な実験であれば、得られたデータや画像、グラフを入力して、『この実験結果はどのように解釈したらよいのですか?』と尋ねると、『こういう解釈ができる』、『こういうことがわかる』といった具合に、自然言語でわかりやすく的確な回答や提案をしてくれます。しかし、最先端研究への適用はまだまだこれから。そのため、現在、化学分野に特化した生成 AI の研究に取り組んでいるところです」。こう語るのは、計算科学やデータ科学を用いた物質・材料の研究を専門とする奈良先端科学技術大学院大学の藤井幹也教授だ。

また、AI 研究を専門とする東京大学の松尾豊・岩澤有祐研究室で、化学分野における生成 AI の適用研究を進めている畠山欽特任准教授は、化学分野において生成 AI が貢献できる領域として、①認識、②分析と予測、③計画の3つを挙げる。実際、OpenAI が2023年3月に公開した ChatGPT の「GPT-4」を使った性能評価を行っている。評価内容と結果は次のとおりだ。

まず、①認識では、化合物の分子構造や物性、特徴などを学習させることで、典型的な化学反応の種類やメカニズム、反応条件、生成物など化学事象に関する認識能力をどれくらい獲得できるかを評価。②分析と予測では、与えられた化合物について分析し、その物性を予測するとともに、予測理由を正しく説明できるかを評価。そして、③計画では、詳細な実験計画の提案が可能かを評価した。

「まず、認識に関しては、2023年時点で GPT-4 には大学院生レベルの認識能力があることを確認しました。また、分析と予測に関しては、GPT-4 に有機化合物の分子構造を与えたところ、その物性をかなり正確に予測し、予測した理由を官能基の有無と関連づけて説明できることを確認しました。学習を重ねることで、仮に未知の分子構造が与えられてもその分子がもつ特性などを予測できることもわかりました。そして、計画に関しては、欲しい収率を得るのに必要な化合物の仕込み量や、反応時間を提案できることが確かめられました」(畠山特任准教授)。

競争力のある独自の生成 AI を構築

一方で、化学分野に特化した生成 AI の実現に向けては、学習データ数が少ないことが大きな課題だと藤井教授は指摘する。「そのため、これまで生成 AI には、公開されている論文や実験データが学習されてきました。しかし、国内外の他の研究グループのデータや構築したモデルを使うことができるというメリットがある反面、競争力のある独自の生成 AI を構築することは困難だと感じました」。

そこで、藤井教授が着目しているのが、「プロセス・インフォマティクス」だ。公開されている論文には、最終的な結果のみが書かれている。「1本の論文を完成させるまでには、数年間に及ぶ試行錯誤や失敗があったはずだ。しかし、その過程やそこで得られた知識やノウハウは記載されていないこともあります。独自性や国際競争力が求められる最先端の研究開発の現場においては、プロセスにまつわる知識やノウハウも重要であるため、生成AIにプロセスを学習させるのが重要だと考えています」(藤井教授)。実際、材料開発の分野では、苦労して良い材料を開発したとしても他社にあっていう間にまねされてしまうことがある。しかし、他社が簡単には製造プロセスを真似できない企業は競争力が強い。

「材料研究における試行錯誤のプロセスを逐一データとして、自前のデータベースに蓄積していきます。そして、ある程度、蓄積された段階で、そのデータを生成AIに学習させるのです。それにより、『このような物性をもつ材料のつくり方を教えて』と質問すれば、生成AIが最適な製造プロセスを答えてくれるようになることが期待できます。もちろん、研究の独自性や競争力を維持するには、モデルもデータもクローズドな環境に置くことが重要です」(藤井教授)。

少ないデータ数で学習させる方法の研究も

加えて、藤井教授は少ないデータ数で、いかに高性能な生成モデルを実現するかの研究も進めている。「十分な学習データが存在しないスパース(希薄な)領域においては、学習、生成、検証、再学習を繰り返すクローズドループを行うことで、データ数を増やし、スパース領域においても新材料の提案が可能な生成AIの構築を目指しています」(藤井教授)。

藤井教授が未知の材料を生成するモデルとして選んだのが、「敵対的生成ネットワーク(GAN: Generative Adversarial Networks)」(図1)だ。GANは組成生成AI、組成識別AI、物性予測AI、材料DBの4つで

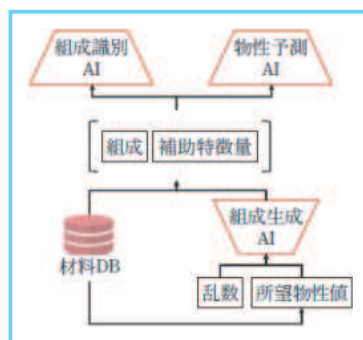


図1 敵対的生成ネットワーク (GAN) の概略図

構成されている。組成識別AIは、乱数などにより組成生成AIが出力した化学組成が、材料DBにある化学組成か、組成生成AIが生成した化学組成かを識別できるように学習するAIだ。一方、組成生成AIは、組成識別AIの正答率が下がるように学習し、あたかも生成した化学組成が、材料DBにある本物の化学組成であるかのように組成識別AIをだます学習をするAIだ。このように、組成生成AIと組成識別AIを競わせることで、組成生成AIはもっともらしい化学組成を生成できるようになる。この一連の学習をクローズドループの中で繰り返すことで、材料DBに収録されているデータ数が少ない領域であっても未知の化学組成を生成する高性能なAIが構築できるというわけだ。

アモルファスや複合材料に関する学習が大きな課題

さて、一概に材料といっても種類がいくつもある。材料の種類によって生成AIのアプローチの方法は異なるのだろうか。「無機材料と有機材料では、AIモデルによる表現方法が大きく異なります。有機材料の場合、分子を文字列に変換しやすい一方、無機材料には、アモルファスの表現のしかたがむずかしいなどの課題があります。金属材料の場合、物性を予測するには、結晶粒同士の界面を考慮する必要がありますし、電池といったデバイスを考える場合は、正極、電解質、負極の複合材料を考える必要があります。これらをどのようにデータ化し生成AIに学習させるかは、今後、取り組んでいかなければならない課題です」(藤井教授)。

画像認識能力が上がった生成 AI の強み

また、化学分野における生成 AI の応用方法として、藤井教授と畠山特任准教授はそれぞれ、実験の自動化に取り組んでいる。生成 AI の画像認識能力が向上していることから、現在、畠山特任准教授は論文などのテキストデータだけでなく、映像を使った学習も進めている。具体的には、実験のプロセスの映像を生成 AI に見せることで、生成 AI に実験内容を認識させる研究だ(図2の一番左)。「どのタイミングで、どのような試薬を何グラム入れ、どのような装置を使って、どのような操作を行ったかなど、一連の実験プロセスを映像で学習させることで、生成 AI に『実験のフローチャートを書いて』と言うと、あっという間に書けるようになりました(図2の左から2番目)。数年以内には生成 AI が実験ノートを書いてくれるようになるでしょう」(畠山特任准教授)。加えて、熟練者による実験テクニックやコツも、生成 AI が学習してデータベースに自動的に蓄積できるようになれば、「研究の大幅な効率化につながる」と畠山特任准教授は語る。

「生成 AI に関しては、分析・予測の領域が期待されがちですが、最初に生成 AI の利便性を実感できるのは、このような人にとって面倒なタスクではないでしょうか。

研究開発の現場で人が行っていることや見ていることをありのままに生成 AI に記録・学習させられるようになってきたことは大変意義深く、大きなブレークスルーであると感じています」と畠山特任准教授は力を込める。目標は実空間から仮想空間までを一気通貫で生成 AI に学習させることだ。

熟練の実験化学者のようなロボットの登場も

加えて、実験に関してはロボティクスとの融合が不可欠だ。すでに、産業分野では、生成 AI を搭載した人型の調理ロボットの開発などが進められており、目の前の食材を使って献立を自律的に考え、料理して客に提供するといったデモンストレーションも披露されている。ロボットによる料理同様に、生成 AI を搭載したロボットが化学実験を自律的に行えるようになれば、実験の大幅な効率化が期待できる。

実際、藤井教授が代表を務めた NEDO「マテリアル革新技術先導研究プログラム」(データ駆動科学が導くスマートスケラブルケミストリ)では、ロボットを使ったポリマー合成の実験に関する研究を行った(写真1)。「これは、奈良先端科学技術大学院大学の情報科学領域の松原崇充教授と佐々木光助教の研究です(arXiv:2506.11384)。本研究では、モーションデモンストレーションとジグ操作の統合によって、化学実

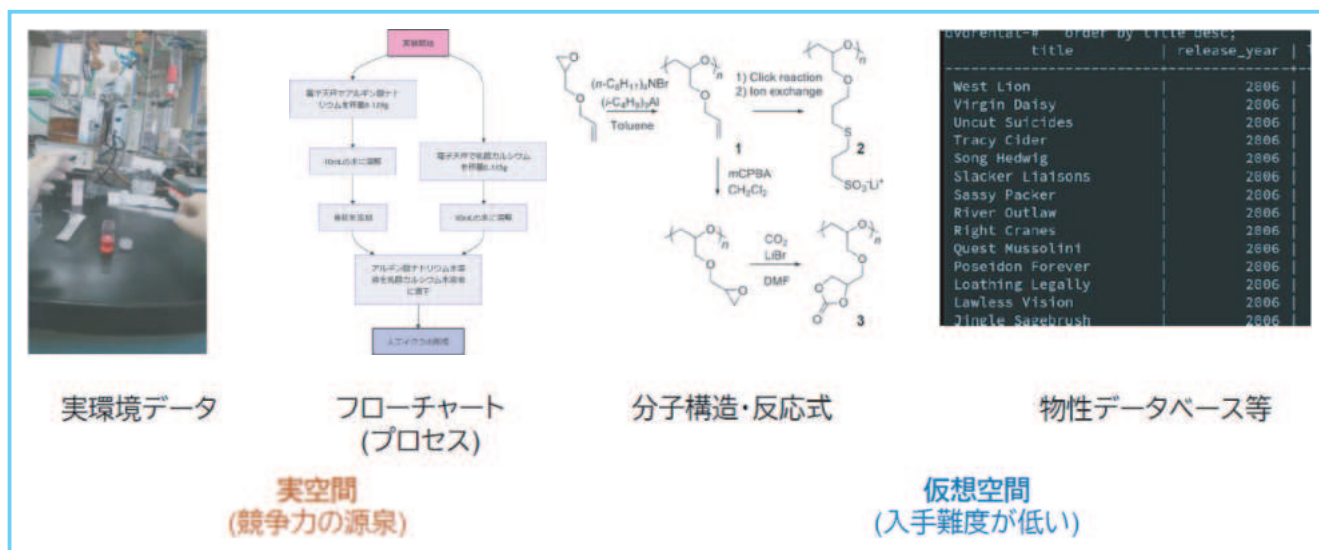


図2 目標：実空間-仮想空間での一気通貫での学習

実験の手技、フローチャート、化学反応、物性データベースなど実空間と仮想空間にまたがる多様な研究データを一気通貫に処理できる生成 AI の開発競争が始まっている。



写真1 ロボットを使ったポリマー実験を行うロボットシステムの概要(松原教授, 佐々木助教からの提供による)

験の柔軟な自動化を可能にする新たなコンセプトが提案されました。実験では、ロボットによる動作の高い再現性とタスクの高い成功率が確認されています。今後は、センサや模倣学習の導入によって、より高い汎用性を持つ自動化システムへの発展が期待されています」(藤井教授)。

近い将来、実験化学者の仕事は生成AIを搭載したロボットに置き換わるのだろうか。「現在のところ、ロボットよりも人の方が手先が器用なので、まずは人が実験を行い、その横で生成AIが指示を出すといった段階を踏むのではないのでしょうか。特に有機材料の場合、ロボットに不向きな作業が多く発生するため、人が行った方が効率的です。その傍らで実験者が間違ったことをしそうになったら、すかさず生成AIが『違いますよ』とダメ出しをするといったイメージですね」(畠山特任准教授)。

一方、藤井教授は次のように指摘する。「実験では、『あれ、このスペクトルはなぜ出ているのだろう』といった小さな気づきが化学の発展につながります。実験における失敗は化学にとって大きな財産です。しかし、生成AIにそのような小さな気づきができるかどうかは現時点では未知数です」。

例えば、生成AIが車の自動運転を行う場合、現時点では交通事故など減多に起こらないレアイベントに関しては学習データが十分でない点が課題となっている。化学分野においても、生成AIが減多に起こらない現象に対応するための方法論はまだ確立していない。加えて、「自動運転による交通事

故同様、化学の実験中に生成AIやロボットが事故を起こした場合、誰が責任を取るのかといった法整備も今後大きな課題になっていく」と藤井教授は語る。

喫緊の課題は人材育成

「生成AIの研究には子育てのような面白さがある」と語る畠山特任准教授。「まだまだ学習能力は低いものの、少しずつ賢くなっていく様子を見守ることに大きな喜びを感じています」。藤井教授も「生成AIを利用しない日はないほど頻繁に利用しています。論文を読む際に不明な点があると、ChatGPTに何度も質問を投げかけます。それにより理解が深まることを実感しています。今後は、生成AIの助けなしに論文を読むことはなくなるのではないかと思うほど効果的です。話し相手にもなってくれるので、生成AIとの会話を楽しんでいます」と語る。

一方で、大きな課題の1つとして、藤井教授も畠山特任准教授も化学と生成AIの両方に精通した人材の育成を挙げる。「アメリカや中国をはじめ各国では、化学の分野で生成AIを活用する動きが活発化しています。アメリカやイギリスの化学会の雑誌にも、頻繁に生成AIを用いた投稿論文が現れていますし、特にアメリカ材料学会は生成AIの話題が非常に多いです」(藤井教授)。「しかし、日本にはそのような人材がほとんどいないのが実情です。私が実験化学者でありながら、AIの研究室に移籍したのも、ひとつには化学と生成AIの橋渡しをして、次世代の化学を担う人材を育成するためです」(畠山特任准教授)。

化学に限らず、物理学や生物学など科学の幅広い分野で急速に存在感を増している生成AI。国際競争力の向上を図っていく上でも科学と生成AIの両方に精通した人材の育成が喫緊な課題といえそうだ。

(サイテック・コミュニケーションズ 山田久美)

OVERVIEWは化工誌編集幹事会の企画・監修により制作されています。