

【環境・省資源分野】

仮訳

化学の言葉を学習する新しいコンピューティングツール（米国） ChemReasoner が大規模言語モデルと深遠な構造的知見を組み合わせ 新たな触媒設計を提案

2024 年 8 月 22 日

著者： [Sarah Wong, PNNL](#)



ChemReasoner は、生成 AI の能力とコンピューターケミストリー（計算機化学）を組み合わせ、新しい触媒設計を支援する。
（イラスト：Cortland Johnson | Pacific Northwest National Laboratory）

化学反応を起こすための、より新しくより優れた触媒開発が試行錯誤的なプロセスを通じて実施される中、[パシフィックノースウエスト国立研究所 \(PNNL\)](#)、イリノイ大学アーバナ・シャンペーン校およびマイクロソフトの科学者らが、このようなプロセスを容易にする [ChemReasoner](#) を開発した。これは、生成人工知能(AI)の力と[計算機化学](#) を組み合わせることで、大規模言語モデル(LLM)に化学の「言葉」を教え、数十年分の化学的知識を迅速に統合し、触媒開発の有望な戦略の特定を支援するものである。筆頭著者で PNNL のデータサイエンティストの Henry Sprueill 氏は、[7 月 25 日にオーストリアのウィーンで開催された国際機械学習会議](#)と、8 月 8 日にカナダのバンクーバーで開催された 2024 Accelerate Conference でこの研究成果を発表している。

[触媒](#)は、[プラスチックのアップサイクル](#)から[持続可能な航空燃料\(SAF\)](#)の生産まで、多くの産業に変革をもたらしてきた。ChemReasoner の最も重要な目標は、新しい触媒材料とエネルギー効率の高い化学反応プロセスの発見を可能にすることである。

「この目標を達成するには、特定の化学反応を選択的に実行できる材料を設計する必要があります」と Sprueill 氏は言う。「つまり、そのような材料の構造に取り込む要素について詳細な判断を下す必要があるということです、これは非常に難しい問題です」。

化学の言葉を学ぶ

1 年半前にこの開発プロジェクトが始まった当時、研究チームは科学者らが日常的に使用している実験プロセスの再現を試みていた。

「実験を設計する際の課題を理解するために、まず触媒研究者らと話しをすることから始めました」と、[ChemReasoner の主任設計者である Sutanay Choudhury 氏](#)は言う。

Choudhury 氏はこうした対話を通じ、PNNL の触媒専門家である Mariefel Olarte 氏と Udishnu Sanyal 氏が科学文献を徹底的に調べて新しい触媒に関する情報を特定し、それらの新しい触媒が目的の化学反応を起こせるかどうかを調査する実験を行っていることを知った。

その後、Choudhury 氏は[コンピューターサイエンティストの Khushbu Agarwal 氏](#)と Sprueill 氏を採用して科学文献検索プロセスを自動化し、それを大規模言語モデル (LLM) の形で実験知識に組み合わせた。ただし、課題は生成 LLM が触媒研究者らと同じ「言葉」を話さないことであった。

「商用 LLM は、新しい触媒構造を提案したり、好ましい経路に反応を誘導したりできるほど化学を理解していません」と Choudhury 氏は言う。「このプロジェクトを開始した当時、ChatGPT 等の LLM のアウトプットは科学百科事典の内容に似ていて、その論理的判断は次世代触媒の発見に必要な深みに欠けているという科学者からの反発が数多くありました」。

触媒、LLM、グラフニューラルネットワーク、計算機化学の専門家を含む ChemReasoner チームは、ここで諦めず、この課題をチャンスと捉えた。

シミュレーションによる出力の強化

LLM は、出版された文献から追加的な補助なく情報を引き出すことができるが、そのアウトプットの精度は大きく変わる可能性があり、完全に創造することさえもある。

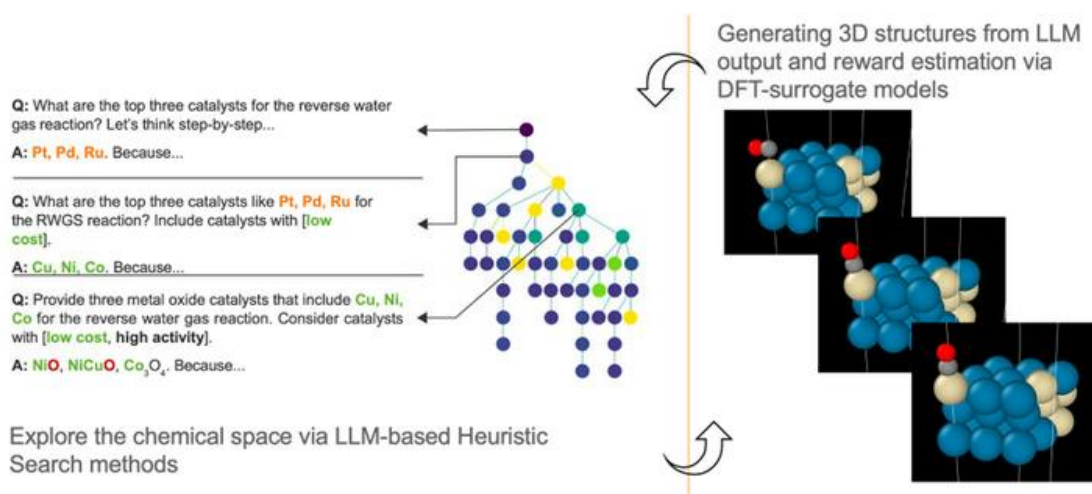
「化学における AI の真の課題は、科学的データによるこれらのモデルの訓練が非常に制限されていることです」と、[PNNL Battelle Fellow](#) で化学者の [Johannes Lerche 氏](#) は言う。「触媒ではこのことがさらに大きな課題となります。最も高収率な方法を用いたとしても、他の科学分野に比べて生成されるデータが少ないのです」。

ChemReasoner 開発チームは、LLM の化学的な論証能力の不足を補う必要があった。ChatGPT 等の一般的な LLM は、人間によるインプットに基づいてアウトプットを変えるように設計されている。これは、「人間のフィードバックによる強化学習(RLHF)」と呼ばれる原理である。

「そこで、『シミュレーションフィードバックによる学習』システムを発明したらどうか、というアイデアが生まれました」と Choudhury 氏は説明する。

ChemReasoner 開発チームは、言語モデルが新しい設計を提案した後、その精度をチェックする手段として量子化学シミュレーションからフィードバックを取得できるよう ChemReasoner を設計した。LLM を[化学シミュレーションデータで訓練されたグラフニューラルネットワーク](#)と組み合わせ、機械学習フィードバックループを作成した。このツールは、ユーザーの定義するパラメータに適合する様々な触媒を評価する手段として、シミュレーションで計算された吸着エネルギーを利用する。ChemReasoner 開発チームは、触媒クエリにおいて、現在利用可能な最先端の LLM である GTP-4 よりも ChemReasoner が優れていることを確認している。

この結果は、吸着エネルギーや反応エネルギー障壁などの触媒固有の概念を取り入れることで、AI システムをエネルギー的に有利な高効率の触媒に導くことを示している。



ChemReasoner は特異性、コストや活性等、ユーザーのニーズに合わせた多種の触媒の調査を可能にする。

(画像提供: Sutanay Choudhury | Pacific Northwest National Laboratory)

自律的な科学に向けて

ChemReasoner はすでに有望な成果を上げているが、開発チームはその能力を拡張する計画を立てている。現在は、CO₂ のメタノールへの変換において、ChemReasoner が選んだ触媒の効率性を実験的に検証する作業に取り組んでいる。

ChemReasoner は、PNNL Laboratory Directed Research and Development program の Generative AI initiative とマイクロソフトの Accelerate Foundation Models Research initiative、そして Azure Quantum Elements とのパートナーシップにより支援された。

訳：NEDO（担当 イノベーション戦略センター）

出典：本資料は、パシフィックノースウェスト国立研究所(PNNL)の記事“Novel Computing Tool Learns the Language of Chemistry”

(<https://www.pnnl.gov/publications/novel-computing-tool-learns-language-chemistry>) を翻訳したものである。