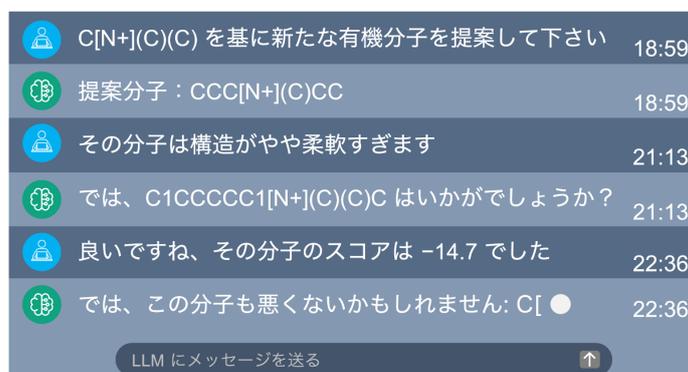


大規模言語モデルを用いた有機分子設計手法の開発

——AI と対話して分子を設計——

発表のポイント

- ◆ 大規模言語モデルを用いた有機分子設計手法を開発した。
- ◆ 蓄積された経験的知識と分子シミュレーションの結果を自然言語に書き起こし、大規模言語モデルに与え続けることで効果的な分子を提案した。
- ◆ 時間とコストのかかる実験による試行錯誤や高度な背景知識が要求される分子設計を、人間と言語モデルの共創によって効率化することに成功した。



大規模言語モデルを用いた対話的な分子設計

概要

東京大学大学院工学系研究科の伊東 周昌大学院生、村岡 恒輝助教、中山 哲教授は、大規模言語モデル (LLM) による有機分子設計手法を開発しました。蓄積された科学的な知識とシミュレーション結果を自然言語を介して活用することで、LLM を有機分子設計に活用できることを示しました。

有機分子を目的の応用に適した形に設計することは複雑なタスクであり、経験を積んだ実験化学者による試行錯誤が必要です。そのため、コンピュータを用いた大量の情報処理が可能になった現代でも、新しい有機分子の開発は実験化学が先導しています。

そこで、実験化学の経験知と、コンピュータの情報処理能力を組み合わせた共創的な有機分子設計を実現するために、LLM を活用しました。LLM は大量のテキストデータで訓練された機械学習モデルであり、さまざまな分野のトピックをカバーしたテキストを人間のように生成できます。LLM が自然言語で人間と対話できる特性を活用し、分子シミュレーションによって生成されたビッグデータと、長年にわたる分子設計の知見を自然言語で LLM に教えることで、新しい有機分子の候補を得ることに成功しました。これにより、実験により性能が実証されている有機分子、それに類似した有機分子、さらには完全に新規でありながら有望な候補を生成することが可能となりました。このアプローチは、分子設計のみならず、医薬品開発やその他の複雑な化合物設計にも新たな可能性をもたらすと考えられます。

発表内容

有機分子の設計は、創薬や環境問題の解決など、人類が直面する課題に対処するための基本的な手法です。有機分子を設計するためには、無数の出発試薬の組み合わせからなる多次元で広大な「化学空間」の中から、合成可能性、安定性、目的物性といった複数のパラメータを同時に満たす候補を割り出す必要があります。従来の分子設計は、主に実験化学者が熟練の経験に基づいて高コストな試行錯誤を重ねることで進められてきました。しかし近年では、ビッグデータを活用したコンピュータに基づく分子設計手法が開発され、実験では扱いにくい多数の分子についても設計・評価することが可能になりました。しかし、確率的なアルゴリズムによって提案される分子は、研究者の直感に反するものも多く、時には合成不能な分子を提案することもあるため、人間の知見は依然として重要です。

大規模言語モデル (LLM) は、膨大なインターネット上のテキストに対して訓練された機械学習モデルであり、人間のような自然な文章生成が可能です。さまざまな分野の質問に答え、議論が行えることから、人間と機械が協働するための有力なプラットフォームといえます。チャットサービスとして有名な訓練済み大規模言語モデルは、翻訳や質問応答、創造的なコンテンツ生成など多岐にわたるタスクで高い性能を発揮しており、一般的な対話から専門的な議論まで柔軟に対応可能です。そのため、分子設計のタスクにおいても、人間と機械が相互にフィードバックを行いながら効果的に協働する可能性が期待されるため、本研究の着想に至りました。

本研究で開発したシステム (図 1) では、LLM が提案する有機分子に対して、分子シミュレーションの結果や、経験的な知識に基づくフィードバックを自然言語で与えることで、分子を改善していくことができます。

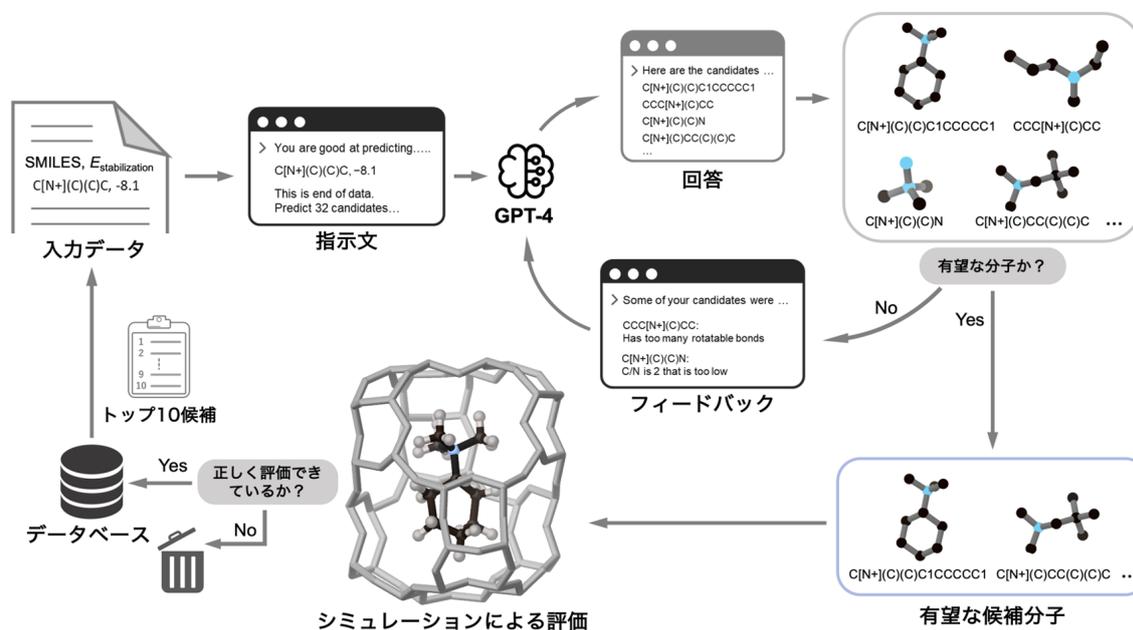


図 1: LLM を用いた分子設計システムの概要

一例として化学工業における重要物質である無機多孔質材料「ゼオライト」の結晶化を促進する有機物 (有機構造規定剤) の設計を試みました。有機構造規定剤は通常、四級アンモニウムカチオンの構造をとり、ゼオライトの結晶構造内部の空間に入り込むことで、特定の結晶相の生成を促進します。ゼオライトの工業的な製造においても広く利用される有機構造規定剤で

すが、製造コストの大部分を占めるため、より低コストな代替分子の探索も大きな課題となっていました。本研究で開発した LLM を用いたシステムを適用したところ、ゼオライトの内部の形状に特異的にフィットするような有機物を設計できました。これにより、実験により性能が実証されている有機分子や、それに類似した有機分子、さらには完全に新規でありながら有望な候補を生成することに成功しました (図 2)。このアプローチは、分子設計のみならず、医薬品開発やその他の複雑な化合物設計にも新たな可能性をもたらすと考えられます。

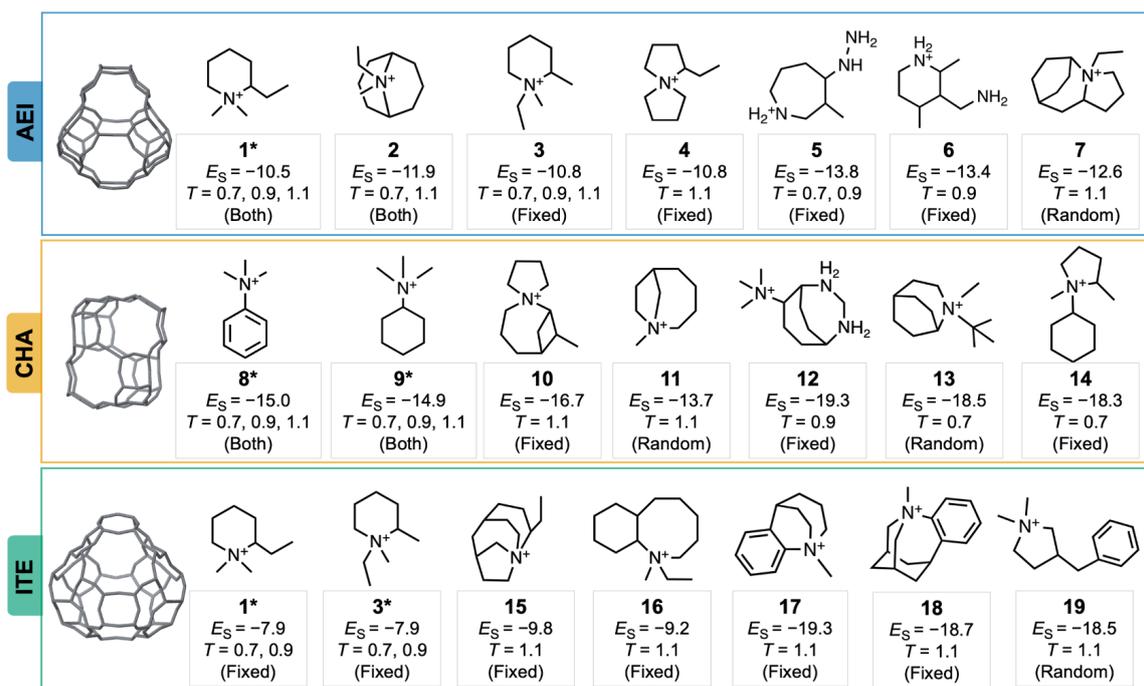


図 2 : AEI 型ゼオライト、CHA 型ゼオライト、ITE 型ゼオライトに対して、LLM が提案した有機構造規定剤 (1-19)

発表者・研究者等情報

東京大学大学院工学系研究科 化学システム工学専攻

伊東 周昌 修士課程

村岡 恒輝 助教

中山 哲 教授

論文情報

雑誌名 : Chemistry of Materials

題名 : Knowledge-Informed Molecular Design for Zeolite Synthesis Using General-Purpose Pretrained Large Language Models Toward Human-Machine Collaboration

著者名 : Shusuke Ito, Koki Muraoka*, Akira Nakayama*

DOI : 10.1021/acs.chemmater.4c02726

URL : <https://pubs.acs.org/doi/10.1021/acs.chemmater.4c02726>

研究助成

本研究は、科研費「準安定結晶性材料の合成可能領域予測モデルの開発(課題番号:22K14751)」、JST さきがけ「地球環境と調和しうる物質変換の基盤科学の創成 原子シミュレーションによるゼオライト触媒のデータ駆動的設計(課題番号:JPMJPR2378)」の支援により実施されました。

問合せ先

(研究内容については発表者にお問合せください)

東京大学大学院工学系研究科
教授 中山 哲 (なかやま あきら)

東京大学大学院工学系研究科 広報室