

幅広い化学反応に対応する AI フレームワーク「CatDRX」

－ 高性能触媒設計を加速し、持続可能な化学・製薬産業に貢献 －

【ポイント】

- 化学反応条件に則した触媒を提案する生成 AI モデル CatDRX を開発
- 多様な触媒反応に対する触媒活性の予測と、新規触媒候補物質の提案を両立
- 生成 AI により、化学産業における有用触媒開発の加速に期待

【概要】

東京科学大学（Science Tokyo）情報理工学院 情報工学系の大上雅史准教授、Kengkanna Apakorn（ケンカーンナー・アーパーコーン）大学院生、菊池雄太特任助教と、九州大学 大学院薬学研究院の丹羽節教授は、効率的に触媒設計を可能にする新たな生成 AI 手法を開発しました（図 1）。

高性能な触媒（用語 1）の設計は、化学反応に必要な時間やエネルギーといったコストを削減できる他、反応過程で生じる廃棄物も抑制できるため、あらゆる化学産業において達成すべき目標となっています。しかし、従来の触媒開発には数年を要することが多く、近年では開発初期段階における触媒候補の選定を効率化するため、機械学習手法の活用が試みられてきました。ただし、これまでのアプローチは特定の反応に特化しており、より汎用的に適用可能な手法の開発が求められていました。

大上准教授らは、生成結果に条件を指定できる条件付き変分オートエンコーダー（CVAE、用語 2）に着目しました。CVAE に基づき、触媒反応における各種反応条件（反応物、生成物、試薬、反応時間など）と、触媒構造から個別に特徴量を抽出・学習し、指定した反応条件に対応できる触媒を提案する CatDRX（Catalyst Discovery framework based on a ReaXion-conditioned variational autoencoder）という事前学習済み生成モデル（用語 3）を開発しました。この手法により、任意の化学反応を促進する触媒の構造を提案すると同時に、その触媒活性を予測することが可能となりました。

この研究成果は 2025 年 10 月 23 日（現地時間）に英科学誌「*Communications Chemistry*」でオンライン公開されました。



図1 CatDRXによって入力した反応条件に最適な触媒構造が提案される枠組みのイメージ図

●背景

化学産業においては、合成プロセスにかかるコストの削減や廃棄物の最小化が強く求められています。プロセスにおいて、触媒の性能は反応に必要なエネルギーや選択性を大きく左右するため、上記の課題を解決する上では高性能な触媒の開発が不可欠です [参考文献1]。しかし、ほぼ無限ともいえる候補の中から、経験則や実験ベースの試行錯誤だけで特定の反応に最適な触媒を見つけ出すことは容易ではありません [参考文献2]。そのため、計算機を活用して触媒を効率的に設計する手法の確立が、触媒開発の加速に向けた重要な鍵となります。

これまでにも、既存の実験データを活用して新規触媒を設計する機械学習や AI 手法が数多く提案されてきました [参考文献3]。しかし、既存手法の多くは限られた反応条件に特化するように設計されており、幅広い反応に対応可能な汎用的手法の開発が依然として求められています。

●研究成果

本研究では、新たな触媒の候補を提案すると同時にその触媒の性能を予測する AI 手法の開発にあたり、指定した条件に基づく生成結果が得られる条件付き変分オートエンコーダー (CVAE) に着目しました。実在する触媒反応から、反応条件と触媒構造を統合的に学習することで、指定した反応条件に対して望ましい特性を持つ触媒を提案できます。また、反応条件はさらに反応物、生成物、試薬、反応時間などの要素に分解し、それぞれ個別に**特徴量** (用語4) を抽出します (図2)。このように個別に抽出した特徴を統合した触媒反応ベクトルを CVAE に学習させることで、さまざまな条件に適した触媒を柔軟に提案できるようになりました。

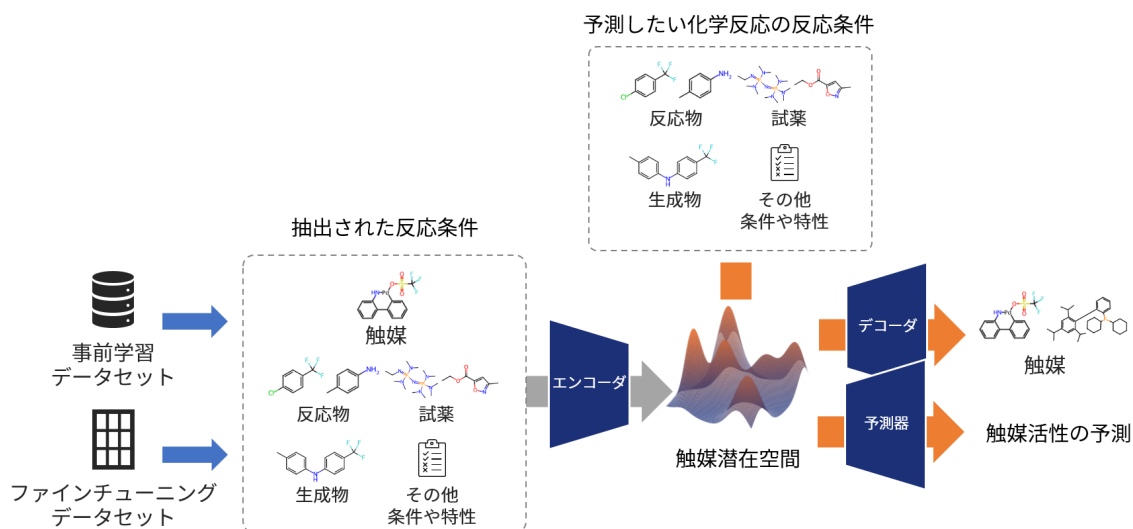


図2 提案手法である CatDRX の全体像

触媒構造や種々の反応条件の特徴から触媒潜在空間を学習し、触媒の提案と性能予測を行う。

本研究で構築した事前学習済み生成モデル「CatDRX」は、以下に示す触媒埋め込みモジュール、条件埋め込みモジュール、およびオートエンコーダーモジュールの3つのモジュールで構成されます（図3）。

- 触媒埋め込みモジュール：触媒構造を表現した**マトリックス**（用語5）から**ニューラルネットワーク**（用語6）により特徴ベクトルを抽出する。
- 条件埋め込みモジュール：化学構造を分子**グラフ**（用語7）として表現し、**グラフニューラルネットワーク**（用語8）で特徴ベクトルを抽出する。また、その他の反応条件は直接的な数値特徴量と**ワンホットベクトル**（用語9）を結合したベクトルとして扱う。
- オートエンコーダーモジュール：これらの特徴ベクトルを統合的に学習することで、指定された反応条件に適した触媒を提案できる。また、オートエンコーダーモジュールには反応収率の**予測機構**が組み込まれている他、必要に応じて他の触媒活性を予測する機構も個別に追加可能である。

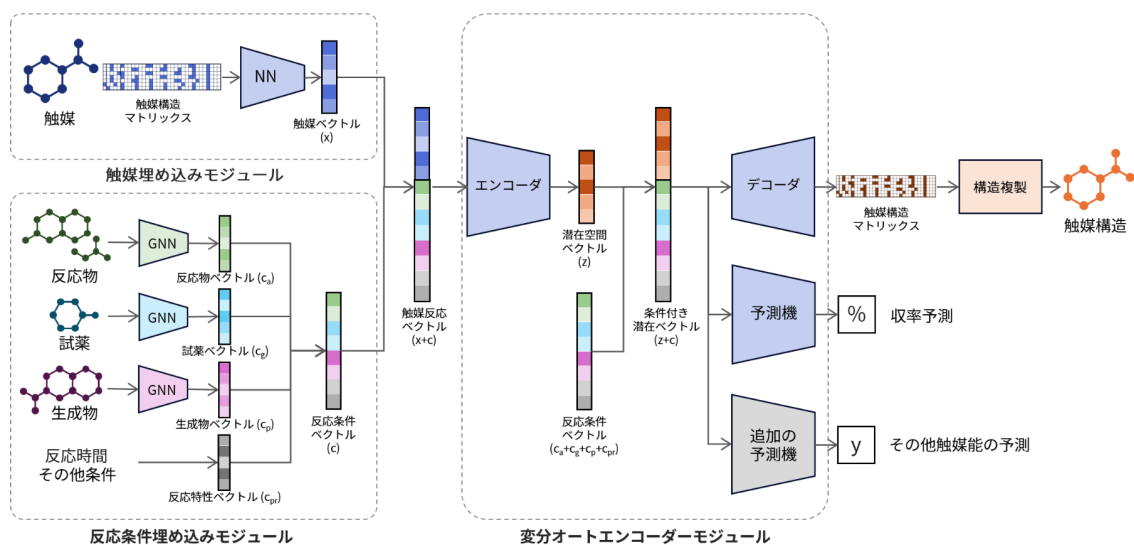


図 3 CatDRX のデータ処理構造

CatDRX では入力データの触媒、反応条件（反応物、試薬、生成物、時間、その他の条件）を個別に特徴量ベクトルに変換した後に統合的に学習する。反応条件を与えることで、任意の反応に適した触媒の構造を提案すると同時に、その性能を予測する。

触媒の提案および触媒活性予測タスクにおける CatDRX の性能を評価した結果、本手法は広範な反応条件に対して安定して良好な性能を示しました。4 種類の異なる反応条件を入力して触媒を提案させたところ、反応それぞれでよく見られる触媒の概形を保持しつつ、多様な構造を提案できることが示されました。また、**潜在空間**（用語 10）からのサンプリングを既存触媒に近い分布から行うか、完全ランダムに行うかを使い分けることで、提案する触媒の妥当性と探索範囲のバランスを調整可能にしました（図 4）。さらに、8 種類の異なる化学反応における触媒活性予測の比較でも、CatDRX は既存手法と同等もしくはそれ以上の性能を示しました（図 5）。一方で、C-C クロスカップリングのように、予測したい化学反応や触媒が学習データと乖離している場合には、予測性能が十分に発揮されず性能が低下するという課題が見られました。これは AI 予測における一般的な課題でもあり、学習に用いるデータや学習対象とする特徴を拡充することでより広範な反応系への適用可能性が高まると考えられます。

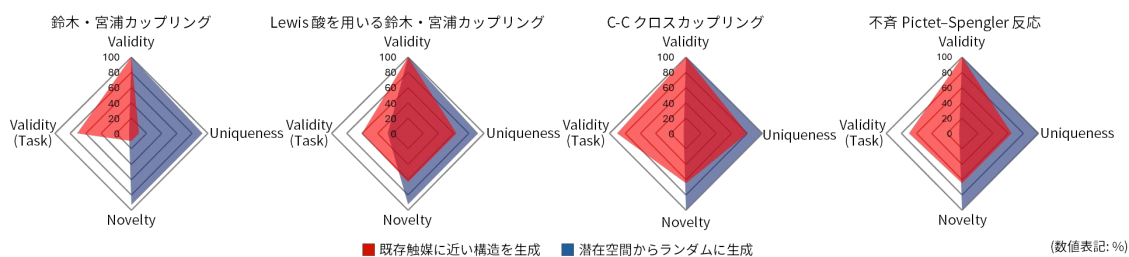


図 4 触媒提案性能の評価

4 種類の反応に対して提案させた触媒の Validity、Uniqueness、Novelty、Validity (Task) を算出した。Validity は構造が化学的に矛盾しないものの割合、Uniqueness は重複せずユニークであるものの割合、Novelty は学習データに同じ構造が含まれないものの割合、Validity (Task) は反応に対して有効と判断されたものの割合を表す。指定した化学反応に有効な触媒を得たい場合は、既存触媒に近い構造から提案させる方法が効率的である。一方、多様な構造の触媒を得たい場合は、潜在空間からランダムに提案させる方法が有効であると言える。

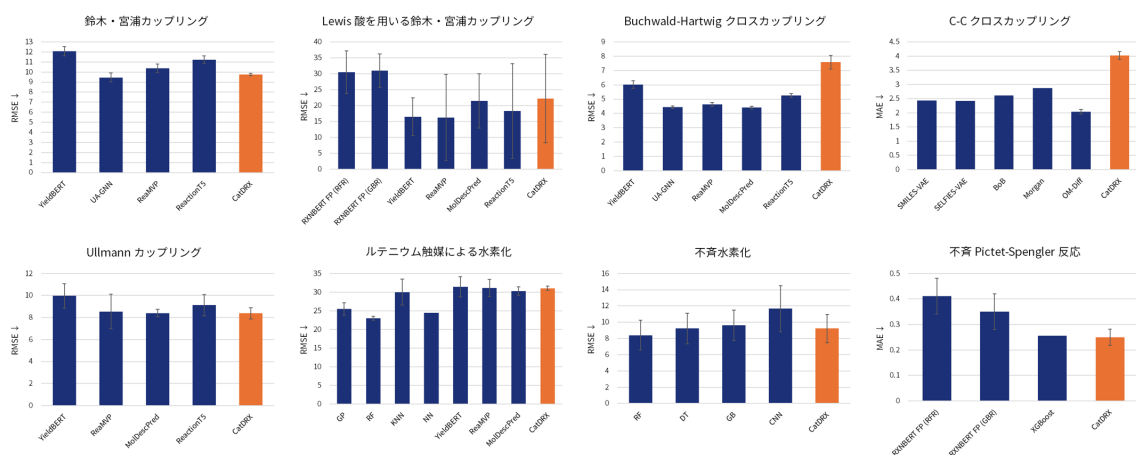


図 5 触媒活性予測性能の評価

8 種類の反応について、実際に使用される触媒の性能を複数モデルで予測し、その結果と実験値との誤差を RMSE (二乗平均平方誤差) または MAE (平均絶対誤差) で評価した。得られた誤差の平均値を棒グラフで示し、エラーバーには標準誤差を用いた。なお、RMSE と MAE はいずれも値が低いほど性能が高いことを示す。

●社会的インパクト

今回発表した CatDRX は、オープンソースソフトウェアとしてプログラム共有サイト GitHub からダウンロード可能です (<https://github.com/ohuelab/CatDRX>)。本研究成果は AI と人間が協力して高度な成果を目指すヒューマン・イン・ザ・ループ (用語 11) の考え方を加速するものと言えます。また、本成果は、情報科学を専門とする大上准教授と有機合成化学を専門とする丹羽教授が共同して取り組むことで達成された、分野横断型の研究成果であり、複雑化・多様化が進む現代の研究分野において、異なる専門性を融合することの重要性を示す好例と言えます。

●今後の展開

研究グループは、事前学習に用いるデータセットを拡張することで、より多様な反応や触媒に対応できる汎用的なモデル構築を目指しています。また、触媒活性予測や提案される触媒の実現可能性のように、一部のタスクでは新たな課題がみられ、今後更なる改善を検討しています。近年、SDGsの観点から化学産業においても環境に配慮した技術が求められており、本研究の成果や今後の発展が、持続可能な社会の実現に向けた一助となることを期待します。

●付記

本研究は以下の事業の支援を受けて実施されました。

- 日本学術振興会 (JSPS) 科学研究費助成事業 学術変革領域研究 (A) 「天然物が織り成す化合物潜在空間が拓く生物活性分子デザイン」 (JP23H04880)
- 日本学術振興会 (JSPS) 科学研究費助成事業 学術変革領域研究 (A) 「bRO5 化合物の潜在空間構築と応用のための情報科学」 (JP23H04887)
- 日本学術振興会 (JSPS) 科学研究費助成事業 学術変革領域研究 (A) 「複雑分子の化学選択的改変による化合物資源の効率展開」 (JP23H04890)
- 科学技術振興機構 (JST) 創発的研究支援事業 「マルチモダリティ創薬を拓くインフォマティクス基盤」 (JPMJFR216J)
- 日本医療研究開発機構 (AMED) 生命科学・創薬研究支援基盤事業 (BINDS) (JP25ama121026)

【参考文献】

- [1] Waclawek, S., Padil, V.V., Černík, M.: Major advances and challenges in heterogeneous catalysis for environmental applications: a review. *Ecological Chemistry and Engineering* **25** (1), 9 (2018)
- [2] Yang, W., Fidelis, T.T., Sun, W.-H.: Machine learning in catalysis, from proposal to practicing. *ACS Omega* **5** (1), 83–88 (2019)
- [3] Moñan-González, L., Burnage, A.L., Nova, A., Balcells, D.: AI Approaches to Homogeneous Catalysis with Transition Metal Complexes. *ACS Catalysis* **15** (11), 9089–9105 (2025)

【用語説明】

- (1) **触媒**：反応開始に必要な活性化エネルギーを低下させることで化学反応を促進する物質のこと。触媒自身は反応の前後で化学的に変化しない。また、選択性を付与できるため、望ましい反応経路を優先させて副生成物を抑制できる。

- (2) **条件付き変分オートエンコーダー (CVAE)**：入力データから潜在空間を確率分布として学習する変分オートエンコーダーに条件を与えて出力結果を制御できるようにした生成モデルのこと。例えば、犬の画像を学習した変分オートエンコーダーは学習で得た確率分布に従ってランダムな犬の画像を生成するが、条件付き変分オートエンコーダーでは指定された犬種の画像を生成できる。
- (3) **事前学習済み生成モデル**：大規模なデータセットであらかじめ学習され、データの特徴や分布を獲得している生成モデルのこと。ファインチューニング（小規模な追加の学習）を行うことで、新しいタスクや条件に対する性能を向上させられる。
- (4) **特徴量**：対象の性質や属性といった特徴を、数値で表現したもの。
- (5) **マトリックス**：数値（ここでは特徴量）を縦横に規則的に並べた表形式データのことを指す。
- (6) **ニューラルネットワーク**：神経回路にみられるニューロンの接続を模した多層構造を持ち、入力データを各層で段階的に処理することで複雑な特徴を自動的に学習するモデルを指す。
- (7) **グラフ**：鉄道の路線図のような、点や丸で表された「ノード」と、それらの間に張られる「エッジ」によって表現される数学的な構造。化合物のグラフ表現においては、原子を「ノード」、原子間の結合を「エッジ」とした原子グラフが通常よく用いられる。
- (8) **グラフニューラルネットワーク**：グラフ情報を扱うニューラルネットワークを指す。
- (9) **ワンホットベクトル**：ある対象を表すために、その対象に対応する成分を 1 とし、それ以外の成分を 0 として表現したベクトルを指す。
- (10) **潜在空間**：学習が進むことで構築される、入力データ群の本質的な特徴で表現される抽象的な空間を指す。生成モデルは潜在空間からサンプリングした潜在ベクトル（= 潜在空間上の 1 点）から新しいデータを生成する。
- (11) **ヒューマン・イン・ザ・ループ**：AI など自動化が行われているシステムなどで、一部の判断や制御に人間を介在させる仕組み。

【論文情報】

掲載誌：*Communications Chemistry*

論文タイトル：Reaction-conditioned generative model for catalyst design and optimization with CatDRX

著者：Apakorn Kengkanna, Yuta Kikuchi, Niwa Takashi, Masahito Ohue

DOI：10.1038/s42004-025-01732-7

【研究者プロフィール】

大上 雅史（オオウエ マサヒト） Masahito OHUE
東京科学大学 情報理工学院 情報工学系 准教授
研究分野：AI 創薬、計算化学

ケンカーンナー・アーパーコーン Apakorn KENGGANNA
東京科学大学 情報理工学院 情報工学系 大学院生（博士後期課程）
研究分野：ケモインフォマティクス、AI

菊池 雄太（キクチ ユウタ） Yuta KIKUCHI
東京科学大学 情報理工学院 情報工学系 特任助教
研究分野：天然物化学

丹羽 節（ニワ タカシ） Takashi NIWA
九州大学 大学院薬学研究院 精密分子変換化学分野 教授
研究分野：有機合成化学

【お問い合わせ先】

（研究に関すること）
東京科学大学 情報理工学院 情報工学系 准教授
大上 雅史
Email：info@li.comp.isct.ac.jp
TEL：045-924-5530 FAX：045-924-5523

九州大学 大学院薬学研究院 精密分子変換化学分野 教授
丹羽 節
Email：niwa@phar.kyushu-u.ac.jp
TEL：092-642-6877

（報道取材申し込み先）
東京科学大学 総務企画部 広報課
取材申し込みフォーム：<https://forms.office.com/r/F3shqsN7zY>
Email：media@adm.isct.ac.jp
TEL：03-5734-2975 FAX：03-5734-3661



九州大学 広報課

Email : koho@jimu.kyushu-u.ac.jp

TEL : 092-802-2130