



触媒シーズ創出に向けた自動特徴量設計技術を開発

～事前知識なしで未知材料の機能を高精度に予測～

ポイント

- 機械学習を用いた材料の機能予測において、経験的な側面を排除した特徴量設計技術を開発
- 事前知識を必要とせず、様々な触媒系のスモールデータに対して圧倒的な予測精度
- 機械学習を用いた材料探索の裾野を大きく広げ、材料シーズ創出を飛躍的に効率化

北陸先端科学技術大学院大学物質化学フロンティア研究領域の谷池 俊明 教授らは、北海道大学大学院理学研究院の高橋 啓介 教授らと共同で、機械学習を用いた材料の機能予測において、事前知識^[注1]を必要とすることなく高精度な予測を実現する、特徴量設計技術を開発しました。

最近、AI やその他の機械学習技術を利用して、触媒などの実用材料に関する研究開発を加速させる取り組みが注目されています。これには、機械を訓練するためのデータと、材料を記述し機能を予測するための変数（記述子）が必要です。中でも、未知材料の機能を高精度に予測するには、機能に影響する因子を効率的かつ網羅的に取り入れた、材料記述子^[注2]の存在が必要不可欠です。従来、この記述子は、対象に関する高度な専門知識（事前知識）に基づいて研究者が手ずから設計してきました。しかし、これは裏を返せば、真新しい、ないしは、非常に複雑などの事由により、知識の蓄積が十分でない対象に対しては、本来最も望まれるにも関わらず、機械学習の活用には大きな制限がありました。

本研究では、対象に対する事前知識を一切必要とせず、数十点程度の訓練データに対して機能する汎用的な特徴量設計技術を開発しました。これは、考え得る大量の記述子候補、すなわち仮説を生成し、目的にかなった記述子を機械に選ばせる、いわば仮説スクリーニング技術です。本研究では、この開発技術が、対象とする触媒反応によらず、従来技術を圧倒する予測精度を与えることや、ハイスループット実験^[注3]と再帰的に組み合わせることで、膨大な候補材料から多様なシーズをピンポイントで見つけられることを示しました。本研究の成果は、機械学習を用いた材料探索の裾野を大きく拡大し、材料シーズ創出の飛躍的な効率化に役立つことが期待されます。

本研究成果は、2024年1月12日10時（英国時間）に英国の科学誌「Communications Chemistry」のオンライン版で公開されました。

本研究は、科学技術振興機構(JST)「未来社会創造事業 探索加速型(No.JPMJMI22G4)」、
「戦略的創造研究推進事業 CREST(No.JPMJCR17P2)」の支援を受けたものです。

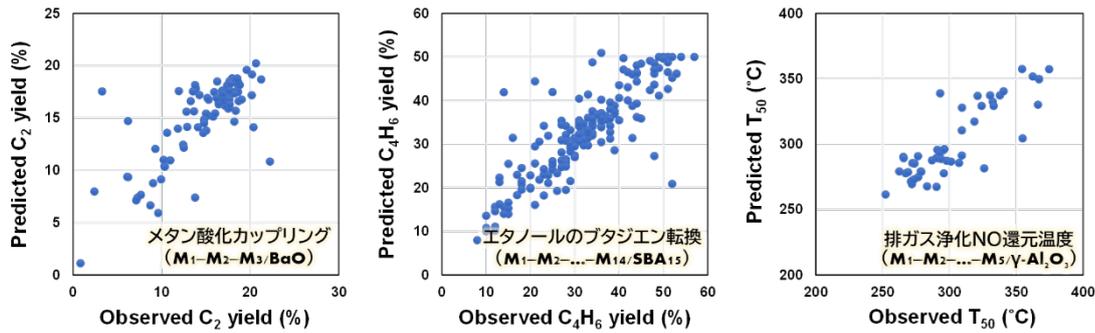
<研究の背景と経緯>

従来、自然科学研究は、個々の研究者の洞察に基づく仮説検証に導かれてきました。しかし、データ駆動型アプローチの隆盛により、このパラダイムは変化しつつあり、触媒を含む様々な材料分野で成功を収め始めています。このような背景の中、効果的な機械学習に適した、十分な規模を備えた材料データの欠如が大きな課題となっています。データの限界は、精巧な機械学習モデルの応用を困難にし、それでも高精度な予測を得るためには、材料の本質を捉えた記述子の存在が必要不可欠です。しかし、このような記述子設計は、関連要因を網羅するために、対象材料の高度な事前知識を必要とし、一般的に大変困難です。何より、未知の領域に踏み込むためにその事前知識が必要となることは論理的に矛盾しており、この記述子設計の経験的な側面は、データ駆動型アプローチの適用対象を、比較的良く知られた材料系に限定せざるを得ない主要因となってきました。

<研究の内容>

今回、本研究グループは、対象材料の事前知識を必要とせず、効果的な記述子を自動的に設計可能な汎用技術を開発しました。開発技術は、材料データが小規模であることを前提とし、元素などに関する一般的な物理量から演算を通して大量の記述子候補を生成し、目的に関連する記述子を機械に選択させる技術です。記述子候補を、材料の機能を説明し得る“仮説”と捉えると、開発技術は、コンピュータ上で大量の仮説を生成し重要な仮説を抽出する、いわば、仮説スクリーニング技術です。本研究では、メタンの酸化カップリング、エタノールのブタジエンへの転換、三元触媒のライトオフ温度という全く異なる対象に対して、開発技術が、触媒組成を記述子とする従来法と比較して、はるかに優れた予測精度を与えることを明らかにしました(図1)。さらに、ハイスループット実験と当該技術を組み合わせる能動学習^[注4]を通じて、機械が触媒設計を捉える認識の精度と汎化能力を改善していき(図2)、最終的に、類似性の低い多様な高性能触媒を、83%もの高精度^[注5]でピンポイント予測することに成功しました。

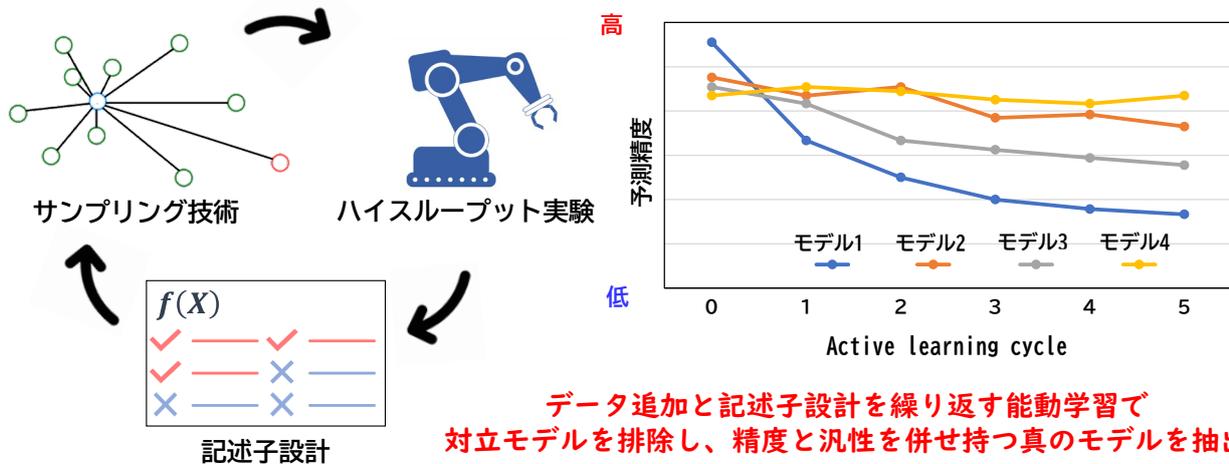
このような成功の裏には、谷池教授らのグループが開発したハイスループット実験技術と、これによって創出した高品質な触媒データの存在が不可欠でした。



5,568個から8個の記述子を選抜、訓練データを差し替えるだけで汎用的に高精度な回帰



図1 開発技術を用いた触媒性能の予測。(上部) 開発技術が異なる触媒反応に対して高精度な回帰を与えること、(下部) 従来技術(元素組成のみ, 元素組成+特徴量選択, 特徴量付与のみ)と比較して極めて高い精度を与えることが示されています。1~3次は合成特徴量の次数を指し、次数が増加するほど、より複雑な特徴量をより大量に生成します。



データ追加と記述子設計を繰り返す能動学習で
対立モデルを排除し、精度と汎性を併せ持つ真のモデルを抽出

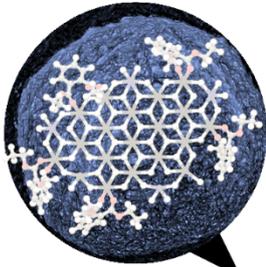
図2 能動学習を用いた機械の改善。汎性を有さないモデル(対立仮説)の予測精度は能動学習に伴い悪化するが、汎性を有するモデル(真の仮説)の予測精度は悪化しない。

<今後の展開>

開発技術は汎用性が高く、触媒に限らず、訓練データを差し替えるだけで様々な材料対象へ即座に展開可能です。本研究グループは、開発技術とハイスループット実験、計画的なサンプリングを組み合わせることで、数十億種もの材料を含むような極めて広大な空間から、事前知識や仮定を一切必要とすることなく、効率的に材料シーズを発見することができるようになって考えています。今後は、開発技術をソフトウェア化し、広く社会実装していく予定です。

<参考図>

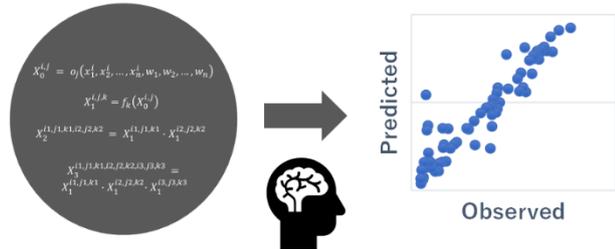
従来:事前知識に基づく記述子設計
新規/複雑な材料系で高難度化...



組成、構造、機構...?



開発技術:自動特徴量設計
対象の事前知識を必要とせず
スモールデータに対して機能する汎用技術



コンピュータ上で大量の記述子候補 = 仮説を
生み出しスクリーニング

自動特徴量設計技術:

AI や機械学習を用いた材料機能の予測において、材料機能を説明し得る材料の特徴、すなわち材料記述子の質は機械学習の精度に直結します。今回開発した汎用技術は、材料記述子の設計を自動化・非専任化する技術です。対象の事前知識を必要とせず、数十点のデータから高精度な学習を可能にします。

<用語解説>

注1) 事前知識

特定材料系の構造や機能などに関する専門知識を指す。従来のデータ駆動型アプローチでは、事前知識に基づき材料記述子を設計することが専らであった。よって、事前知識が十分に存在しない、複雑ないし未知の材料系では材料記述子の設計が困難であった。

注2) 材料記述子

組成や構造、物理特性といった材料を特徴付ける量の中で、目的とする材料機能と関連するものを材料記述子と呼ぶ。材料機能が単一の材料記述子によって説明できることは稀である。一方、材料記述子の数を増やすほど規模の大きなデータが求められるため、データが小さくなりがちな材料分野においては、機能を十全に説明可能な最小数の記述子を入力することが肝要である。

注3) ハイスループット実験

自動化・並列化・効率化などの手段に基づき単位時間当たりの実験数を飛躍的に増大させた実験を指す。材料分野では、研究者間にデータ取得・報告に関する統一性が存在しないことがほとんどであり、均質なデータを効率的に生成できるハイスループット実験は、データ駆動型アプローチと親和性が高い。

注4) 能動学習

データを追加し繰り返し学習させることで機械の精度や汎化能力を高める方法を指す。

注5) 83%もの高精度

予測した触媒（36種）の内、エタンとエチレンの合計収率が15%以上を達成した触媒（30種）の比率を示す。

<論文情報>

雑誌名：Communications Chemistry

論文タイトル：“Automatic feature engineering for catalyst design using small data without prior knowledge of target catalysis”

（対象の事前知識を必要としない触媒設計のための自動特徴量設計技術）

著者：Toshiaki Taniike*、Aya Fujiwara、Sunao Nakanowatari、Fernando García-Escobar、Keisuke Takahashi

doi：10.1038/s42004-023-01086-y

<お問い合わせ先>

<研究に関すること>

谷池 俊明（タニイケ トシアキ）

北陸先端科学技術大学院大学 物質化学フロンティア研究領域 教授

Tel：0761-51-1630 Fax：0761-51-1635

E-mail：taniike@jaist.ac.jp

< J S T の事業に関する事 >

幸本 和明 (コウモト カズアキ)

科学技術振興機構 未来創造研究開発推進部

Tel : 03-6272-4004 Fax : 03-6268-9412

E-mail : kaikaku_mirai@jst.go.jp

< 報道担当 >

北陸先端科学技術大学院大学 広報室

Tel : 0761-51-1031

E-mail : kouhou@ml.jaist.ac.jp

北海道大学 社会共創部 広報課 広報・渉外担当

Tel : 011-706-2610

E-mail : jp-press@general.hokudai.ac.jp

科学技術振興機構 広報課

Tel : 03-5214-8404

E-mail : jstkoho@jst.go.jp