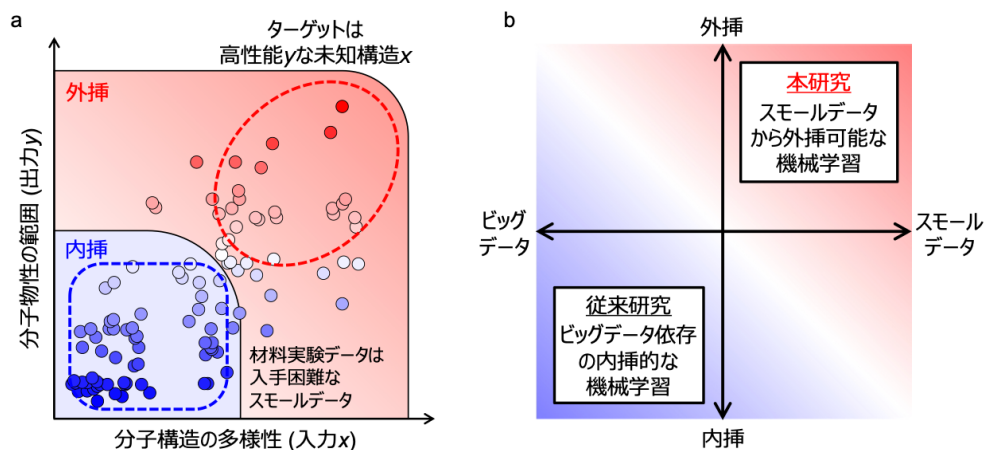


未踏材料の創出に向けて外挿可能な物性予測モデルを開発 ——AIの常識を覆し、スモールデータから外挿的予測を実現——

発表のポイント

- ◆ 機械学習と量子化学計算を統合した物性予測モデルを開発し、大規模なベンチマーク評価により提案モデルが外挿的な材料探索に最適であることを実証しました。
- ◆ 物理に立脚した機械学習モデルにより、ビッグデータに頼るAIの常識を覆し、スモールな実験データのみから材料探索に不可欠な外挿的予測を実現しました。
- ◆ 本研究はマテリアルズ・インフォマティクス課題であった外挿的な材料探索を可能にし、既存材料を凌駕する未踏材料の創出に貢献すると期待されます。



(a) 材料探索に不可欠な物性予測の外挿性

(b) データ規模と外挿性に着目した本研究と従来研究の違い

概要

東京大学大学院工学系研究科電気系工学専攻の嶋川肇大学院生、熊田亜紀子教授、佐藤正寛准教授らの研究グループは、新材料の創出に向けて機械学習と量子化学計算（注1）を統合することで、ビッグデータに頼るAIの常識を覆し、スモールな実験データのみから外挿（注2）的な物性予測を実現しました。提案モデルは、分子構造を量子化学計算で得られる物理量に変換し、さらに分子のカテゴリカルな特徴にもとづいて量子化学情報と物性間の複雑な関係性を抽出します。本研究では、種々の物性予測モデルの外挿性能を評価するため、有機化合物の12種類の実験データを利用した大規模なベンチマークテストを実施しました。その結果、提案モデルが物性予測の内挿・外挿性能ともに優れており、材料探索に最適なモデルであることを実証しました。本研究はマテリアルズ・インフォマティクス（注3）の課題であった外挿的な材料探索を可能にし、既存材料を凌駕する未踏材料の創出に貢献すると期待されます。

本成果は1月10日付でNature系科学雑誌の『npj Computational Materials』オンライン版に掲載されました。

発表内容

計算科学やAIを活用して材料開発を高効率化する手法として、マテリアルズ・インフォマティクスが注目されています。機械学習を用いて材料構造と物性の関係の学習・予測を行い、人間では発見できなかった高機能な新規材料を開発することを目的としています。しかしながら、材料物性の予測を担う機械学習はそもそもデータの内挿を前提とする手法であり、外挿に適していないことが問題でした。新規材料を扱うためには構造情報に対する外挿性が必要であり、より高い機能性を予測するためには、物性値の出力範囲に対する外挿性が必要になります（図1a）。このとき、材料物性の実験データは500点以下のスモールデータであることが多く、ビッグデータ解析を基本とする従来のAI技術では材料探索での適用範囲に限られるため、外挿的な物性予測モデルは未開発のままでした（図1b）。

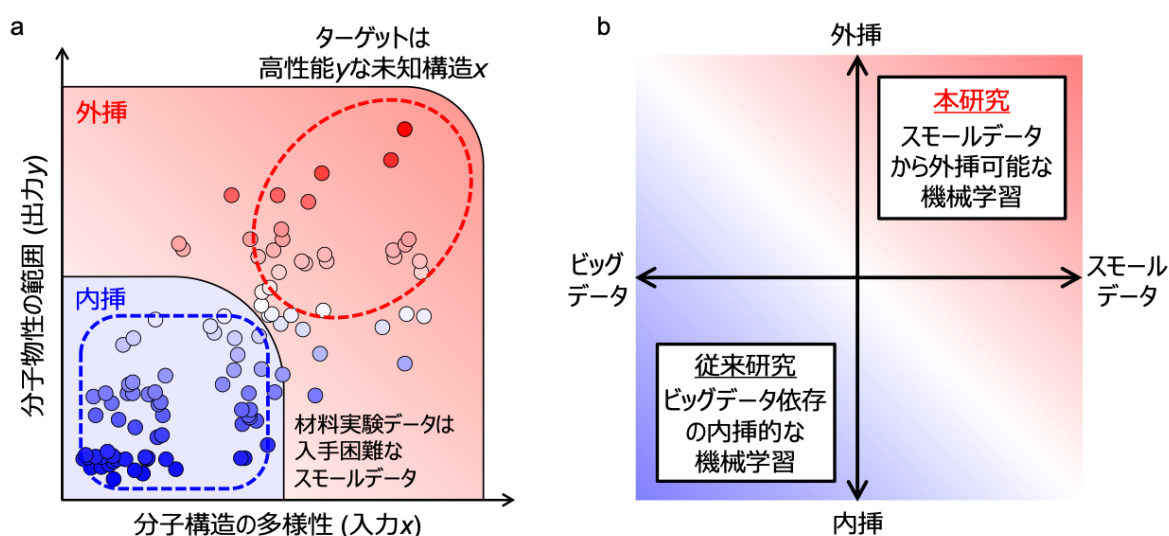


図1：(a) 材料探索に不可欠な物性予測の外挿性
(b) データ規模と外挿性に着目した本研究と従来研究の違い

本研究ではスモールデータのみから外挿的な材料探索を実現するため、機械学習と量子化学計算を統合した物性予測モデルを開発しました。100点から12,000点の有機化合物で構成される12種類の物性実験データを用いて内挿・外挿性能の大規模なベンチマークテストを行い、提案モデルの優位性を実証しました。まず、従来の分子構造ベースの予測モデルは学習データが小規模な場合に内挿性能が不足することに加えて、分子構造または物性範囲の分布が学習/テスト間でシフトするデータに対して外挿性能が欠如することが明らかになりました。そこで分子物性の外挿的な予測を実現するため、量子化学計算で得られる物理量の新たなデータセット、および量子化学情報にもとづく物性予測モデルを提案しました。その中でも特に、量子化学情報と化学カテゴリ情報（分子特徴のバイナリ表現）との相互作用項を用いた物性予測モデル（図2）は、分子構造および物性範囲の外挿テストにおいて最高性能を達成しました。実験データのスモール性および分子構造に対する外挿性は、任意の分子構造に対して変換可能かつ多様な分子物性に対応可能な量子化学情報を媒介した機械学習により克服されました。また、物性範囲に対する外挿性は、相互作用を導入することで表現力を拡張した線形回帰モデルにより実現されました。

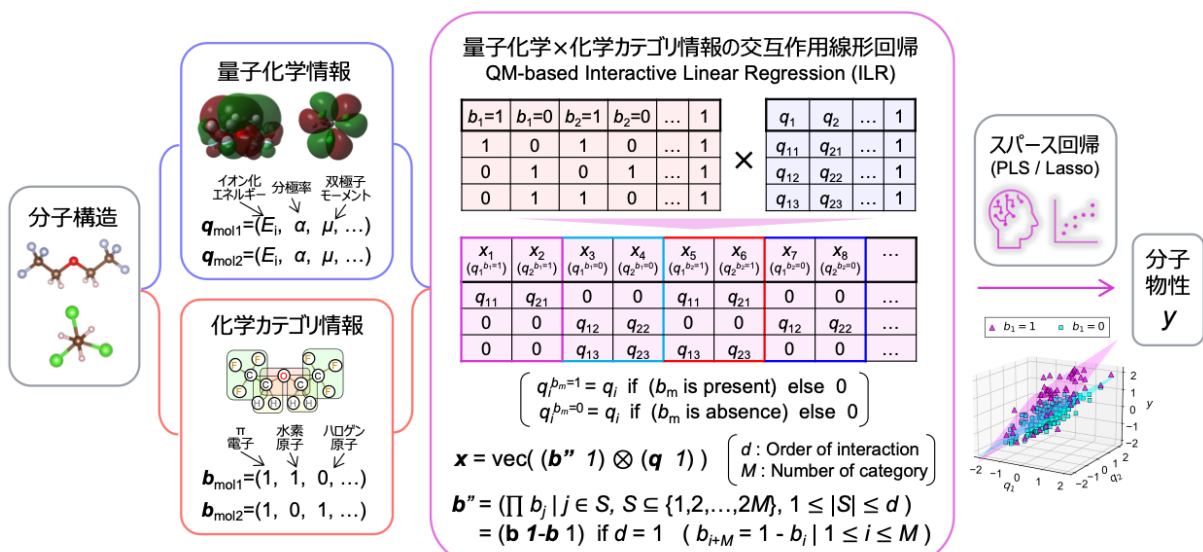


図 2 : 量子化学情報と化学カテゴリ情報を用いた交互作用線形回帰モデル

本研究の成果は、ベンチマークテストの詳細な結果に加えて、約 2.6 万分子に対して実施した量子化学計算の結果や、交互作用項を生成する計算コードとともに公開されています。提案モデルはビッグデータに頼る AI の常識を覆し、スモールな実験データのみから、材料探索に不可欠な外挿的予測を実現します。本成果はベンチマークで扱った材料・物性に限らず一般的な材料探索にも応用可能であり、マテリアルズ・インフォマティクスを活用した多くの分野で既存材料を凌駕する未踏材料の創出に貢献すると期待されます。

発表者・研究者等情報

東京大学大学院工学系研究科 電気系工学専攻

熊田 亜紀子 教授

佐藤 正寛 准教授

嶋川 肇 博士課程

論文情報

雑誌名 : npj Computational Materials

題名 : Extrapolative prediction of small-data molecular property using quantum mechanics-assisted machine learning

著者名 : Hajime Shimakawa*, Akiko Kumada & Masahiro Sato

DOI : 10. 1038/s41524-023-01194-2

URL : <https://doi.org/10.1038/s41524-023-01194-2>

研究助成

本研究の一部は NEDO 未踏チャレンジ 2050、学際大規模情報基盤共同利用・共同研究拠点 (JHPCN) (課題番号: jh230055) の研究助成を受けて行われました。また、本研究の一部は大阪大学サイバーメディアセンターの SQUID を利用して行われました。ここに深く謝意を表します。

用語解説

(注1) 量子化学計算：

分子の構造や性質を第一原理にもとづいて電子状態から非経験的に解析する手法。

(注2) 外挿：

既知である学習データの入出力情報を用いて、学習した範囲外にある未知データに対して行う予測。一方、学習した範囲内にある未知データに対して行う予測を「内挿」と呼ぶ。

(注3) マテリアルズ・インフォマティクス：

計算科学や情報処理技術を活用したデータ駆動型の材料開発手法。機械学習を用いて材料と物性の関係を学習・予測し、所望の機能を持つ新規材料の開発を目的とする。

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 電気系工学専攻
准教授 佐藤 正寛 (さとう まさひろ)

東京大学大学院工学系研究科 広報室