

生成モデルを用いた 固体材料発見手法の追求

東京大学情報基盤センター 華井雅俊
jh250048

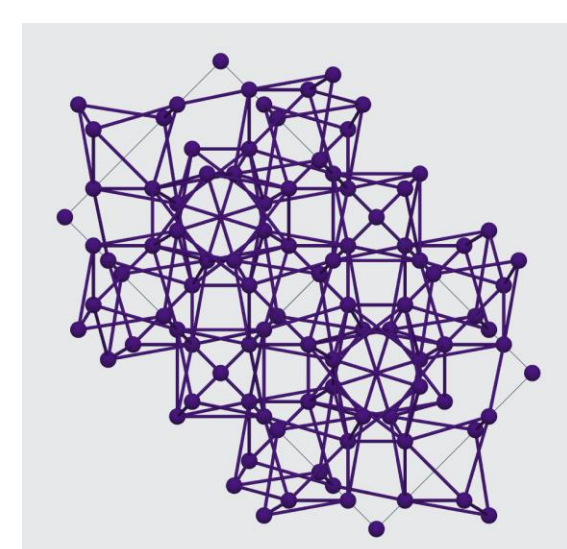


研究概要

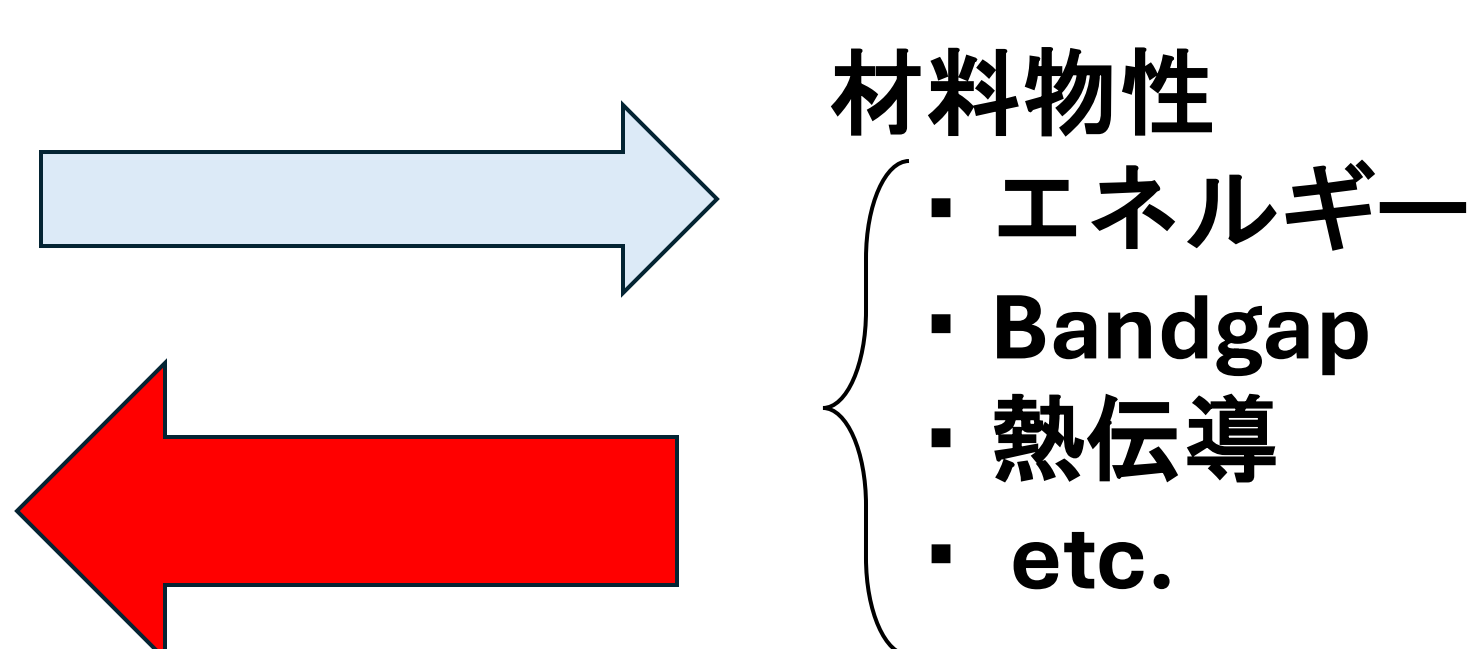
本研究では固体材料の発見的手法の追求を行う。2024年のノーベル物理学賞・化学賞に代表されるように、機械学習手法の自然科学分野における応用が活発であり、材料分野においてもデータ駆動型の材料発見や物性予測の研究が盛んに行われている。機械学習や生成モデルで顕著な成功をおさめる自然言語・画像処理分野と大きく異なり、**自然科学分野、特に材料分野におけるデータ駆動型研究開発の難しさは発見的タスクを本質的に内在することに起因**することが大きい。つまり、既存のデータから未知の材料や特異な物性値を発見すること(外挿)が材料開発では強く求められるが、今日のニューラルネットワークをベースとした機械学習モデルは内挿能力に比べ外挿能力が非常に限定的であることが知られ困難をきわめる。

本研究では根本的な問題に立ち返り **(1) 発見的生成タスクのベンチマーク作成 (2) ベンチマークに基づいた生成モデルの構築および精度追求 (3) 生成モデルの大規模化と実行効率化**を実施することで、発見的な材料向け汎用生成モデルの構築手法を追求する。

結晶の生成問題



結晶構造



材料物性
・エネルギー
・Bandgap
・熱伝導
・etc.

結晶構造(cifデータ)から物性値を予測する順問題、一方、物性値や構成の元素から所望の結晶構造を予測生成する問題を逆問題という。本研究では逆問題を対象とする。特に、**学習データに含まれない新しい構造の発見**を問題の主眼とする。

既存の手法

材料分野における外挿能力の追求はこれまで様々行われる。特に、順問題における外挿予測に関して各提案がある。

・ Transfer Learning・Multi-Task Learningを使った手法

Chanhui Lee et.al. "Scalable Multi-Task Transfer Learning for Molecular Property Prediction" ICMLW AI4Science 2024

・ 外挿の学習方法を学習するMeta Learning手法

Kohei Noda et.al. "Advancing Extrapolative Predictions of Material Properties through Learning to Learn" arXiv. 2024

・ Pretraining + Fine tuningを利用した方法 (成果3)

S. Takashige, M Hanai, et.al "Is Self-Supervised Pretraining Good for Extrapolation in Molecular Property Prediction?" xSIG 2023

しかしながら、未知の構造を発見する**生成タスクにおける外挿問題の研究は未発展**である。

提案

発見的生成タスクのベンチマーク

"未知な物質とはなにか?"という素朴な問いから具体的なベンチマークを設計する

Case 1: 入力による外挿

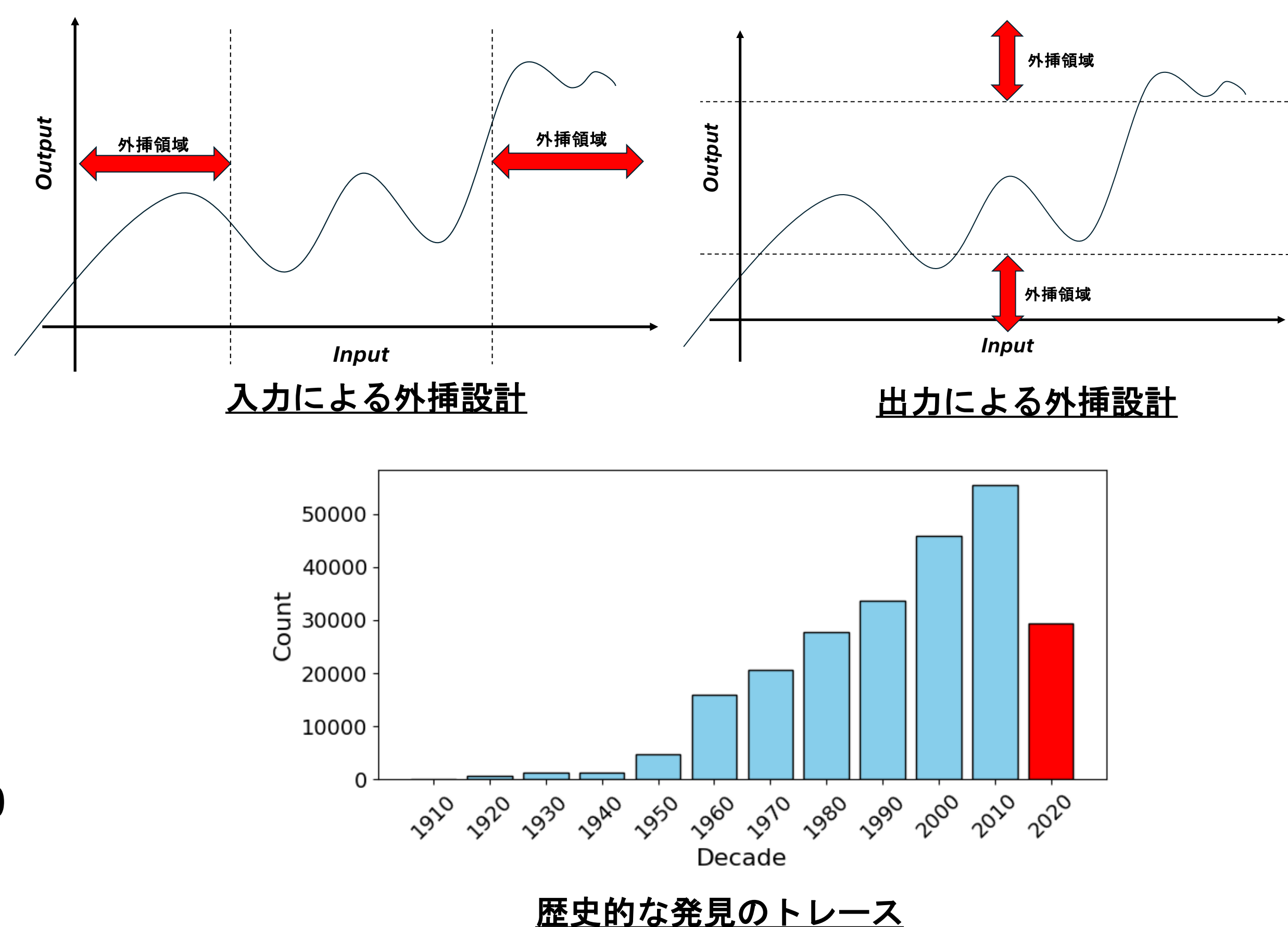
未知の入力に対しても予測モデルが精度を担保できるか? 例えば、未知の構造や未知の元素があった場合の予測能力

Case 2: 出力による外挿

未知の出力に対して予測モデルが精度を担保できるか? 例えば、新しいチャンピオン物性(超伝導や高熱伝導)の予測能力

Case 3: 歴史的発見問題

古い年代のデータから最近の材料を予測できるか? 例えば1900年代の学習データから2000年代の予測を実施



発見的生成モデル

- 既存の外挿物性予測モデルにおける主要なアイデア (Transfer / Multi-task Learning, Meta Learning, Pretraining等) を使った生成モデルの設計
- 生成モデルに特有の外挿向け生成モデル手法の開発

大規模化・高速化

- Deep Speed などLLM向け高速実装の材料タスク向けへの移植とMiyabi上のUnified Memory (CPU + GPU) への性能最適化
- DFT 計算による大規模な固体材料データセットの作成



まとめ・今後

本研究では、これまでの機械学習手法では困難である**“発見的問題”**へ本格的に取り組む。特に固体材料向けの構造生成問題を対象に、ベンチマークの作成、外挿予測問題に特化したモデル構築手法の確立、それらモデルのMiyabiによる大規模化・高速化を実施する。

ベンチマーク作成

発見的生成モデルの構築

大規模化・高速化

