

令和 5 年 6 月 19 日現在

機関番号：62603

研究種目：若手研究

研究期間：2020～2022

課題番号：20K19866

研究課題名（和文）転移学習による外挿的な物性予測：包括的訓練済みモデルライブラリの創出

研究課題名（英文）Transfer learning to predict the properties of compounds: pre-trained model library

研究代表者

劉 暢 (Liu, Chang)

統計数理研究所・ものづくりデータ科学研究センター・特任助教

研究者番号：30814149

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：転移学習は、あるドメインで訓練されたモデルを他のドメインに適用するための機械学習技術で、訓練データが足りないタスクに対してよく使われる訓練手法である。本研究では、転移学習をマテリアルズインフォマティクス(MI)に導入し、材料データ量の少なさ問題の解消と外挿的物性予測の実現を図る。本研究は3年間行い、次の成果を得た：1) 大量の訓練済みモデルを有するモデルライブラリー(XenonPy.MDL)を開発した；2) 結晶系格子熱伝導率とハイエントロピー合金を巡った研究を展開し、研究成果を論文と国際会議で発表した；3) 結晶構造探索の課題に適用し、従来の予測手法の性能を圧倒する予測アルゴリズムを提案した。

研究成果の学術的意義や社会的意義

学術的意義としては、転移学習の技術はデータ収集に高いコストがかかる材料研究分野において必要不可欠である。本研究で挙げられたハイエントロピー合金の熱力学安定性予測と結晶構造予測の研究成果はその実例であり、小規模な第一原理計算の結果のみで高精度な予測を実現した。転移学習技術の導入は、研究の効率化と新たな技術の実現に向けた重要な一歩となった。社会的意義としては、転移学習の導入により、材料の設計や特性予測の精度と外挿性能が向上し、材料開発のスピードが加速されることが期待できる。これにより、エネルギー効率の高い材料や環境負荷の低い製品の開発が促進され、持続可能な社会の実現に寄与することができる。

研究成果の概要（英文）：In this study, we have successfully employed transfer learning, a powerful machine learning technique, to address the challenge of limited material data and enable predictive extrapolation in the field of materials informatics (MI).

Our research has yielded significant outcomes. Firstly, we have developed XenonPy.MDL, an extensive model library containing a multitude of trained models. This library serves as a valuable resource for further advancements in the field. Secondly, we have applied transfer learning on thermodynamic stability prediction of high-entropy alloys and lattice thermal conductivity prediction, leading to notable findings presented at prestigious international conferences. Furthermore, our work has extended into crystal structure prediction, where we have introduced prediction algorithms surpassing the performance of conventional methods. We have achieved remarkable results by proposing an innovative prediction algorithm, particularly in crystal structure prediction.

研究分野：マテリアルズインフォマティクス

キーワード：転移学習 スモールデータ 物性予測 結晶構造予測

1. 研究開始当初の背景

本研究は、マテリアルズインフォマティクス (MI) と呼ばれるデータ科学と物質科学の学際領域に位置付けられる。MI は 2011 年に米国で始動し、日本では、2015 年に国立研究開発法人物質・材料研究機構を拠点に JST イノベーションハブ構築支援事業「情報統合型物質・材料開発イニシアティブ (MI²I)」が始動した。機械学習による物性予測は、仮想スクリーニングへの活用等、材料研究の新たなアプローチとして産学の多方面から注目を集めている。しかし、MI に対する世界的な期待感の高まりと共に、材料データの量的な限界が MI を阻害する要因として顕在化しつつある。スモールデータの問題は、先端領域に近づくにつれて顕著になり、さらに、コミュニティ全体でコモンデータを創出しようという動向もほとんど見えてこない。すなわち、MI の問題点の多くは、データがないということに帰着する。また、機械学習のモデルは、既存のデータとの類似性から未知物質の特性を予測する。したがって、一般には周辺にデータが存在しない真に革新的な物質の特性を予測できない。材料研究の究極の目標は、外挿的予測と発見の実現である。

転移学習は、あるドメインで訓練されたモデルを他のドメインに流用するための解析技術である。例えば、大量の画像を用いて動物の種類を判定するニューラルネットワークを訓練し、少数の花の画像データを用いて訓練済みモデルを改変して花の種類のカテゴリを構築する。動物のカテゴリは、訓練の過程で汎用的な画像特徴量を獲得していることが期待され、その一部は花のカテゴリに転用できる可能性がある。その場合、花のカテゴリを一から学習するのではなく、少数のデータで動物のカテゴリを微修正すれば十分かもしれない。ヒトの脳には、少ない経験でも合理的な予測を実行できる仕組みが備わっている。例えば、幼い頃にピアノを学んでいた人は、音楽に関する一般的な知識を獲得しているため、他の楽器の演奏技術を比較的容易に習得できる。このような推論過程を模倣したものが転移学習である。我々の先行研究により、材料研究の様々なタスク (有機・無機物質の物性予測等) において、転移学習が有する潜在能力、特に転移学習による外挿性の獲得の可能性が明らかになった。本研究では、これらのシーズを深化させるべく、材料研究において転移学習を系統的且つ戦略的に実践していくための理論と方法論及びソフトウェア等の学術基盤を整備し、転移学習を駆動力に材料研究のスモールデータの限界突破を図ると同時にモデルの外挿的予測と言う最大の難問に挑む。

2. 研究の目的

MI 技術の発展は新たな材料の開発やエネルギー効率の向上など、幅広い産業や技術分野において革新的な進展をもたらす可能性がある。転移学習の導入により、材料の設計や特性予測の精度と外挿性能が向上し、材料開発のスピードが加速されることが期待できる。本研究の目的は以下 2 点である。

- ① 転移学習を MI に導入するための理論と方法論及びソフトウェア開発等の学術基盤の構築。
- ② 転移学習を活用した少数データにおいての MI 応用を実践。

3. 研究の方法

機械学習の目的は、系の入力 S に対する出力 Y の予測である。例えば、入力物質 (分子、組成、結晶等)、出力は物性値 (エネルギー等) に相当する。この目的を達成するために、入力を数学的表現する記述子の設計と数学的表現から物性の写像をモデルに学習されるアルゴリズムの開発が必要となる。機械学習の全てのモデルは、入力データのパターン (数学的表現) を読み解き (モデル訓練)、出力を予測する。しかし、MI 問題の複雑さによりモデル訓練用のデータが希少な上、革新的な材料の周辺にもデータは存在しない。このことは、オーソドックスな機械学習では原理的には真に革新的な材料に到達できないことを意味する。転移学習は、データ数の多いドメインで学習した汎用的特徴量 (記述子) をデータ数の少ないドメインに持って来られるため、スモールデータ問題に適切な研究手法と考えられる。また、申請者が東京大学塩見淳一郎教授らとの無機結晶系高熱伝導材料の共同研究で^[1]、転移学習モデルは時に超外挿的予測ができる事実 (図 1) を踏まえて、3 段階の研究を行った。

- ① 訓練済みモデルライブラリ XenonPy.MDL の大幅な拡充を行う (物性・材料項目の増加)。多様な訓練済みモデルの集合体を MI のスキームに組み込むことは、ヒトの脳の情報処理との対比で言えば、多様な経験から獲得し構造化された記憶を推論プロセスに活用することに相当する。訓練済みモデル (=経験から獲得した記憶) のボリュームと多様性の増大こそが、転移学習の潜在能力を引き上げる源泉と考える。
- ② 転移モデルに外挿性を獲得させるための方法論を研究する。現時点では転移モデルの外挿性の有無を事前に判断する方法がない。数理モデルの信頼性評価の方法は、一般に“不確かさの定量” (UQ; uncertainty quantification) と呼ばれるカテゴリで研究が進められてきたが、転移学習の UQ の方法は確立に至っていない。現時点では理論的な根拠は不明であるので、研究構想全体の中で本項目が最も高いハードルになる。
- ③ 実例を通じて、転移学習の能力を実証する。実証に必要な材料合成・実験については、

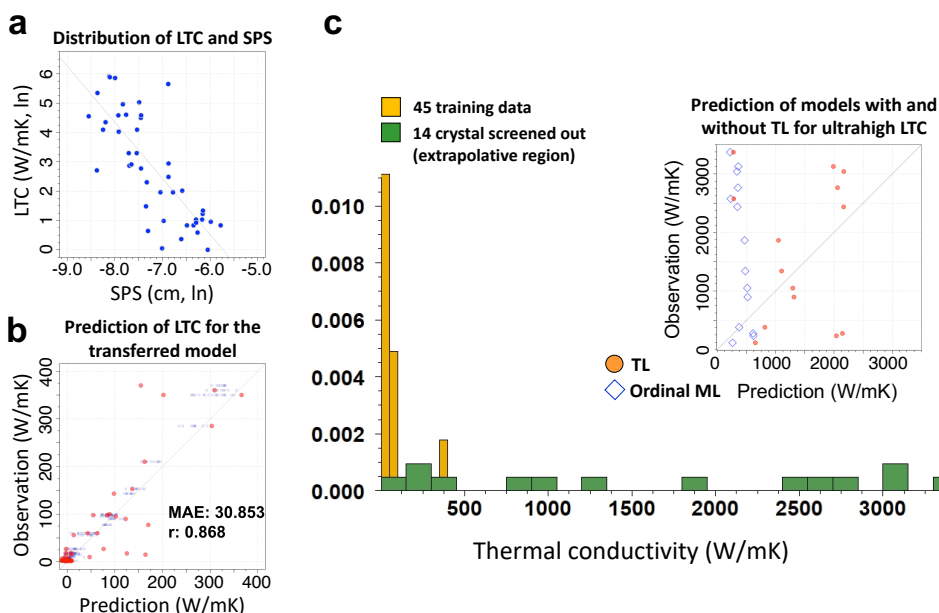


図1. 転移学習による高熱伝導性 (LTC) 無機結晶のスクリーニング. (a) SPS と LTC の相関. (b) 転移学習で得られたモデルの LTC の予測値と実験値. 赤点は検証用データ, 青点は訓練データを表す. (c) 訓練データとスクリーニングで見つかった 14 個の結晶の LTC の分布. 訓練データのほとんどは 100 W/mK 以下の範囲に分布しているのに対し, 14 個の結晶には LTC が 3,000 W/mK を超すものが含まれている. 散布図は, 14 個の結晶に対する LTC の予測値と実験値を表す. 転移学習のモデルの予測精度 (黄) は, 直接訓練されたモデル (青) を遥かに凌駕する. 3,000 W/mK を超すような超外挿領域でも予測精度が保たれていることを確認できる.

4. 研究成果

① 訓練済みモデルライブラリ XenonPy.MDL の拡充及び拡張

XenonPy.MDL には, 低分子・高分子・無機結晶系の 45 種類の物性 (力学的性質, 熱電特性, 光学特性, 局所構造特徴など) に対する約 160,000 個の公開モデルが収録されている. また, ハイエントロピー合金の熱力学的安定性予測と結晶構造探索の実証問題に向けて, 4 万ほどの非公開モデルを訓練した.

本研究の推進と共に, XenonPy.MDL システムに対する要求も大きく変わった. ユーザーのニーズとさまざまな MI 研究のシチュエーションを考え, XenonPy.MDL システムを利用する様々なユーザーがワークフローを自由自在に構築できるように, XenonPy.MDL をクラウドサービスである Amazon Web Services (AWS) に移行する. 現在, 移行作業はまだ完了していないが, 申請者は 2023 年度から実行する研究 (若手研究 23K16955) にも利用される予定である.

② 転移モデルに外挿性を獲得させるための方法論の研究

研究構想全体の中で本項目が最も高いハードルになると想定した. 研究当初と比べて, 外挿的予測の現象は転移学習を適用することで可能になることを明らかにしたが, 理論的な根拠はまだ不明瞭のままである. 現状, 方法論の研究は同じものづくりデータ科学研究センターに所属する特任研究員である南俊匠が受け継ぎ, 初期の研究結果をまとめた論文は現在投稿中である^[2].

③ 転移学習の能力の実証

本研究は結晶系格子熱伝導率予測^[1,3], ハイエントロピー合金の熱力学安定性 (論文制作中) と結晶構造予測問題 (論文投稿中)^[4] に対して, 転移学習の手法を導入した. 抜粋として, 最も重要と考える転移学習を活用した結晶構造予測問題の新規解法について説明する.

結晶構造予測の目標は, 標的化合物の化学式の情報のみから, その化合物が安定に存在する結晶構造を提案することである. すなわち, 化学式から結晶構造を推測し構造エネルギーを評価する. 相対的に低いエネルギーを持つ構造は安定構造と想定されている. 従来の実験手法による結晶構造決定は, 時間とコストがかかるため, 機械学習の発展により, 近年では, サロゲートモデルによる探索手法が数多く提案されたが, 機械学習モデルの予測精度や計算コストの問題など, 様々な課題が依然として存在している.

効率的な結晶構造予測の鍵は、リーズナブルな結晶構造生成器と高精度かつ高速度で構造エネルギーを評価する手法の開発にある。結晶構造生成器について、我々は、教師あり学習で化学式から安定構造のスペースグループ、対称性、類似性などの情報を予測するモデルを訓練し、予測された情報を入力とする、三種類の構造生成器を開発した。構造エネルギーの評価について、我々は、Material Project databaseに保存した結晶構造を入力とするサロゲートモデル（ソースモデル）を訓練し、探索しようとする系に転移学習を行った。第一原理計算で少ないデータでもバーチャル構造に対する予測精度の高いモデルの訓練を実現した（図2）。

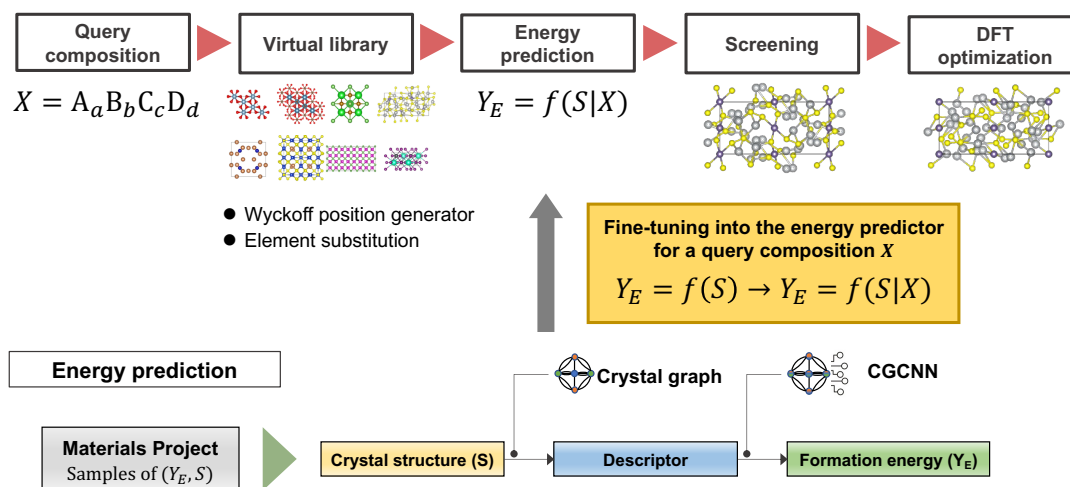


図2. 転移学習を活用した結晶構造探索のワークフロー。

我々は多様性、複雑性、実応用性などの観点から35個のベンチマークを用意し結晶構造予測の性能テストを行った。結果、主流となる構造予測手法の正解率（例えば、USPEX）と比べ、我々の提案手法は最低48.6%、最大91.4%の正確率に達することが分かった（表1）。

表1. 35個のベンチマークにおける正解率の比較。

Composition	Space group-aware	Metric learning based	Template based	USPEX
Ag ₅ GeSe ₆	-	-	✓	-
Al ₂ O ₃	✓	✓	✓	✓
BN	-	-	✓	-
Ba(FeAs) ₂	-	✓	✓	✓
Bi ₂ Te ₃	✓	✓	✓	-
C	✓	-	✓	-
Ca ₁₄ MnSb ₁₁	-	-	✓	-
CaCO ₃	✓	✓	✓	✓
Cd ₃ As ₂	✓	✓	✓	-
CoSb ₃	✓	✓	✓	-
CsPbI ₃	-	✓	✓	-
Cu ₁₂ Sb ₄ S ₁₃	✓	✓	✓	-
Fe ₃ O ₄	-	-	✓	-
GaAs	✓	✓	✓	✓
GeH ₄	-	-	✓	-
La ₂ CuO ₄	-	✓	✓	-
Li ₃ PS ₄	-	-	-	-
Li ₄ Ti ₅ O ₁₂	-	-	✓	-
LiBF ₄	-	-	✓	✓
LiCoO ₂	-	-	✓	-
LiFePO ₄	✓	✓	✓	✓
LiPF ₆	✓	✓	✓	✓
Mn(FeO ₂) ₂	✓	✓	✓	✓
Si	✓	-	✓	✓
Si ₃ N ₄	-	-	✓	✓
SiO ₂	-	-	-	-
SrTiO ₃	✓	✓	✓	-
TiO ₂	-	-	✓	✓
V ₂ O ₅	-	-	✓	-
VO ₂	✓	✓	✓	✓
Y ₃ Al ₅ O ₁₂	✓	✓	✓	-
ZnO	✓	✓	✓	✓
ZnSb	✓	✓	✓	✓
ZrO ₂	-	✓	✓	-
ZrTe ₅	-	✓	✓	-
Overall	17/35 = 48.6%	21/35 = 60%	31/35 = 91.4%	13/35 = 37.1%

<引用文献>

- [1]. S. Ju, R. Yoshida, C. Liu, S. Wu, K. Hongo, T. Tadano, and J. Shiomi, “Exploring diamondlike lattice thermal conductivity crystals via feature-based transfer learning,” *Phys. Rev. Mater.*, vol. 5, no. 5, p. 053801, May 2021, doi: [10.1103/physrevmaterials.5.053801](https://doi.org/10.1103/physrevmaterials.5.053801).
- [2]. S. Minami, K. Fukumizu, Y. Hayashi, and R. Yoshida, “Transfer learning with affine model transformation.” arXiv, Available: [http://arxiv.org/abs/2210.09745](https://arxiv.org/abs/2210.09745)
- [3]. P. Torres, S. Wu, S. Ju, C. Liu, T. Tadano, R. Yoshida, and J. Shiomi, “Descriptors of intrinsic hydrodynamic thermal transport: screening a phonon database in a machine learning approach,” *J. Phys.: Condens. Matter*, vol. 34, no. 13, p. 135702, Mar. 2022, doi: [10.1088/1361-648X/ac49c9](https://doi.org/10.1088/1361-648X/ac49c9).
- [4]. C. Liu, H. Tamaki, T. Yokoyama, K. Wakasugi, S. Yotsuhashi, M. Kusaba, and R. Yoshida, “Shotgun crystal structure prediction using machine-learned formation energies.” arXiv, Available: [http://arxiv.org/abs/2305.02158](https://arxiv.org/abs/2305.02158)

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 4件）

1. 著者名 Liu Chang, Fujita Erina, Katsura Yukari, Inada Yuki, Ishikawa Asuka, Tamura Ryuji, Kimura Kaoru, Yoshida Ryo	4. 巻 33
2. 論文標題 Machine Learning to Predict Quasicrystals from Chemical Compositions	5. 発行年 2021年
3. 雑誌名 Advanced Materials	6. 最初と最後の頁 2102507 ~ 2102507
掲載論文のDOI (デジタルオブジェクト識別子) 10.1002/adma.202102507	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Torres Pol, Wu Stephen, Ju Shenghong, Liu Chang, Tadano Terumasa, Yoshida Ryo, Shiomi Junichiro	4. 巻 34
2. 論文標題 Descriptors of intrinsic hydrodynamic thermal transport: screening a phonon database in a machine learning approach	5. 発行年 2022年
3. 雑誌名 Journal of Physics: Condensed Matter	6. 最初と最後の頁 135702 ~ 135702
掲載論文のDOI (デジタルオブジェクト識別子) 10.1088/1361-648X/ac49c9	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Ju Shenghong, Yoshida Ryo, Liu Chang, Wu Stephen, Hongo Kenta, Tadano Terumasa, Shiomi Junichiro	4. 巻 5
2. 論文標題 Exploring diamondlike lattice thermal conductivity crystals via feature-based transfer learning	5. 発行年 2021年
3. 雑誌名 Physical Review Materials	6. 最初と最後の頁 53801
掲載論文のDOI (デジタルオブジェクト識別子) 10.1103/PhysRevMaterials.5.053801	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Chang Liu, Hiromasa Tamaki, Tomoyasu Yokoyama, Kensuke Wakasugi, Satoshi Yotsuhashi, Minoru Kusaba, Ryo Yoshida	4. 巻 -
2. 論文標題 Shotgun crystal structure prediction using machine-learned formation energies	5. 発行年 2023年
3. 雑誌名 arXiv	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.48550/arXiv.2305.02158	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 2件）

1. 発表者名 Nguyen-Dung Tran
2. 発表標題 First-principles study of Quaternary High Entropy Alloys consisting of Fe-Ni-Co-Cr-Mn/Pd
3. 学会等名 2022 TMS Annual Meeting & Exhibition (国際学会)
4. 発表年 2022年

1. 発表者名 Stephen Wu
2. 発表標題 XenonPy: an open source platform for data-driven materials design with small data
3. 学会等名 TMS 2023 Annual Meeting & Exhibition (国際学会)
4. 発表年 2023年

〔図書〕 計1件

1. 著者名 伊藤 聡、吉田 亮、劉 暢、Stephen Wu、野口 瑤、山田 寛尚、赤木 和人、大林 一平、山下 智樹	4. 発行年 2022年
2. 出版社 共立出版	5. 総ページ数 202
3. 書名 マテリアルズインフォマティクス	

〔産業財産権〕

〔その他〕

XenonPyプロジェクトHP http://xenonpy.readthedocs.io/ XenonPy.MDL検索用ページ（開発中） xenon.ism.ac.jp XenonPyプロジェクトソースコード https://github.com/yoshida-lab/XenonPy
--

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------