

# 研究報告書

## 「機械学習手法による合理的な材料物性予測技術の構築」

研究タイプ: 通常型

研究期間: H27年12月～H31年3月

研究者: 世古敦人

### 1. 研究のねらい

近年、材料科学の諸問題に機械学習手法を応用することで、高精度かつ効率的な材料物性の予測が可能になってきている。その中の一つが第一原理計算の多重実行と機械学習手法を組み合わせた原子間ポテンシャル構築である。この第一原理計算と同程度の精度を持つ原子間ポテンシャルは、第一原理熱力学計算を、実用上重要な多元系材料や欠陥構造、界面などに適用することにおいて、必要不可欠なものである。もう一つが、機械学習手法を使った材料探索である。これは、元素、結晶構造、化学組成など多くの自由度に依存する材料の物性値を、機械学習手法により、効率的に予測することで、膨大な化合物の中からスクリーニングを実施し、所望の材料を探索するものである。

これら2つの研究に共通していることは、予測する物性についてのデータベースをもとに、機械学習手法により物性予測モデルを構築するという点である。このような物性予測モデルにおいては、記述子(物性予測モデル構築のための説明変数)が予測精度の大部分を決定する。その中でも、結晶構造情報をどのように記述子として表すかということは非常に難しい問題である。本研究では、機械学習による物性予測モデルに適した結晶構造を表現する一般的な記述子を考案する。さらに、これらを用いた物性予測モデル構築を実施する。以下の2つの研究項目に対して、考案した結晶構造記述子を応用することにより、結晶構造記述子の有効性や可能性を評価する。

**研究項目Ⅰ 結晶構造記述子を利用した高精度原子間ポテンシャル構築**

**研究項目Ⅱ 元素記述子、結晶構造記述子を利用した物性予測モデル構築手法の開発**

本研究では、2つの研究項目を設定しているものの、物性予測モデル構築手法の基本的な枠組みは同じであるため、両方の研究項目を実施することにより、相乗効果が期待できると考えている。本研究が実現されれば、第一原理熱力学計算の応用範囲が大きく広がるとともに、従来の経験則による材料探索を超えることが可能となることが期待されるなど、材料科学・材料工学におけるインパクトは極めて大きいと予想される。

### 2. 研究成果

#### (1) 概要

本研究では、機械学習のための様々な元素・結晶構造・化合物記述子を導入した。本研究で導入した記述子は、以下の3つに分類される。

①物理・化学・材料科学の分野において、古くから様々な用途で使われている量を拡張し、網羅的な形で記述子として利用したもの。電気陰性度、擬ポテンシャル半径などの元素記述子、動径分布関数、Bond-order parametersなどの構造記述子が該当する。

②元素種や化学組成を直接的に記述子として利用するもの。元素の有無を1または0で表現

する One-hot 表現などが該当する。

③本研究提案の群論的方法およびその拡張により生成可能な化合物データや物性の対称性を満たす系統的な記述子。

また、第一原理計算により様々な物性データセットを構築し、予測モデルを作成することにより、導入した記述子の有効性を実証した。

以下、研究項目ごとの概要を記す。

### 研究項目 I 結晶構造記述子を利用した高精度原子間ポテンシャル構築

〔記述子①〕原子間距離・三体間角度に基づいた構造記述子により、高精度な原子間ポテンシャルを構築する方法を提案した。単体金属 31 種を対象に、高精度なポテンシャルを構築できることを示した。

〔記述子③〕より高精度なポテンシャル作成が可能な系統的構造記述子を生成する群論的手法を構築した。結晶構造に対して網羅的な回転不変量を生成し、原子間ポテンシャルを構築した結果、広範囲の構造に対する予測能力が大幅に向上することがわかった。また、多元系へと手法を拡張し、プロトン配置エネルギーの高精度モデルを構築できることを示した。本手法は、他の一般的な物性にも応用可能なものであると期待される。

### 研究項目 II 元素記述子、結晶構造記述子を利用した物性予測モデル構築手法の開発

〔記述子①③〕元素・局所構造記述子をもとに多様な化合物を統一的に記述するための方法を提案した。凝集エネルギー、融点、バンドギャップ、格子熱伝導率、プロトン伝導の活性化エネルギーのデータに対して、高精度な予測モデルを作成できることを実証した。

〔記述子①②〕無機結晶データベースに推薦システム手法を応用し、合成可能な新規無機化合物を効率的に発見する方法を提案した。100 億以上の化学組成の中から、無機化合物が存在する組成を予測可能にするものであり、新規無機化合物の発見を大幅に加速させることができることを期待される。

## (2) 詳細

### 研究項目 I 結晶構造記述子を利用した高精度原子間ポテンシャル構築

本研究項目では、遷移金属を含む単体金属に対して高精度な機械学習原子間ポテンシャル (MLIP) が構築可能な枠組みを提案した [研究成果 (論文発表) 3, 5 など]。具体的には、データ作成、構造記述子の導入、予測能力評価を段階的に行った。まず、31 種類の単体金属において、多数の結晶構造について第一原理計算を実行し、MLIP 構築のためのエネルギーデータの作成 (約 10 万件) を行った。また、原子間距離および三体間角度に依存する Angular Fourier series を、構造記述子として導入し

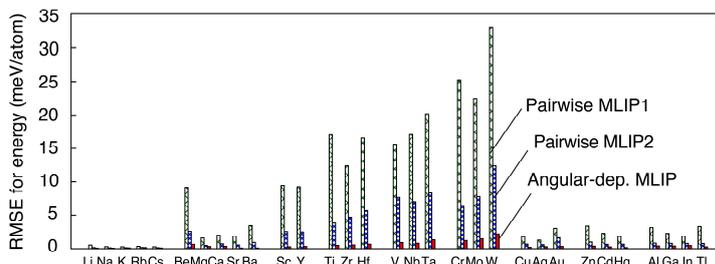


図1 単体金属 31 元素における MLIP の予測誤差。

た. 図1に示すように, 典型金属元素では, 原子間距離のみに依存する構造記述子により高精度な MLIP を構築できた. 一方で, 遷移金属元素では, 三体間記述子が不可欠である. 最終的に, 三体間記述子を考慮した場合の平均予測誤差は 0.9 meV/atom となった. これらの結果は, すべての金属元素において, 統一した枠組みにより高精度な MLIP を構築することができることを示している. 高精度な MLIP を構築すると同時に, MLIP と金属系で広く用いられている EAM ポテンシャルとの関係を明らかにした. これは, ブラックボックス的で物理的解釈が難しい MLIP に対して, 一つの物理的解釈を与えるものである.

また, 研究を進めていく上で, 原子間距離および三体間角度に依存する構造特徴量を用いたモデルを用いた場合, 広範囲の構造に対する予測能力が不十分であることがわかってきた. 導入した原子間距離および三体間角度に依存する記述子は, 動径関数と球面調和関数の積を基底関数として原子分布を展開した場合の1次および2次の回転不変量に相当することに起因する. よって, 群論の手法により網羅的に高次の回転不変量を生成し, MLIP を高精度化する方法を開発した. [論文執筆中]

さらに, 約30の元素種を含むプロトン伝導体のエネルギーデータ(約 60000 件)を用いて, MLIP モデルを多元系へ拡張することにより, プロトン配置エネルギーに対する高精度な MLIP を構築した. その際, 高次の回転不変量を導入することにより予測精度が向上することがわかった[継続中].

この他にも, 第一原理計算の実行が難しい結晶粒界など大規模構造への応用を目指し, 異常検知の機械学習による指標の導入, Semi-supervised learning の応用に取り組んだ. このような方法に基づき, 大規模構造に対する原子シミュレーションを高精度化することが不可欠であるという結論が得られた. [論文執筆中]

このように, 高精度な MLIP が構築可能な系統的・網羅的な構造記述子およびその生成方法を導入できたという観点から, 本研究項目の目的は達成できたと考えている.

## 研究項目 II 元素記述子, 結晶構造記述子を利用した物性予測モデル構築手法の開発

### 1. 多様な化合物を統一的に記述可能な化合物記述子による物性予測モデル

本研究項目では, 多様な化合物を統一的に記述するため, 単位胞中の原子数や化学組成によらない化合物記述子を生成する方法を提案した[研究成果(論文発表)4]. 具体的には, 化合物中の原子を, その原子種を表現する元素記述子や配位環境を表現する構造記述子へと変換し, 元素・構造記述子空間における原子分布の不変量を化合物記述子として導入する方法である.

図2は, 構築した凝集エネルギーのデータ(18096 化合物)を用いて, カーネルリッジ回帰予測モデルを構築し, 予測誤差を評価したものである. 最も良い予測モデルの予測

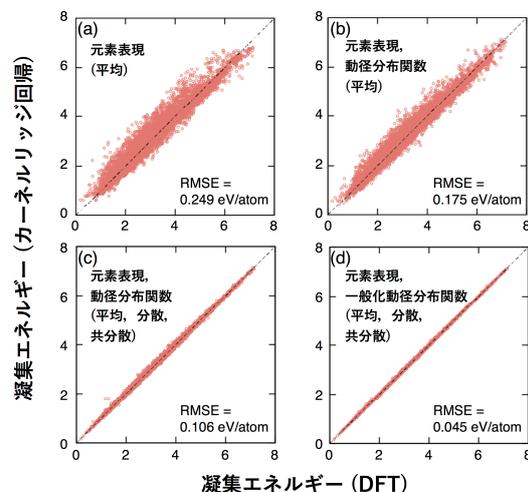


図2 凝集エネルギーの第一原理計算値およびカーネルリッジ回帰予測値.

誤差は、0.045 eV/atom[図2(d)]であり、第一原理計算を実施することなく、1 kcal/mol (0.043 eV/atom)程度のエネルギー予測ができることを意味する。また、適切な元素・結晶構造記述子を導入するだけでなく、化合物記述子の構築方法が重要であることがわかった[図2]。

また、凝集エネルギーと同様に、格子熱伝導率や融点、半導体バンドギャップ[J. Lee, A. Seko et al. Phys. Rev. B 93, 115104 (2016)]についても、高精度なモデルが得られた。また、ガウス過程に基づいたベイズ最適化により、プロトン伝導体におけるプロトンの最安定位置および鞍点位置を効率的に探索する方法を提案した [K. Toyoura et al. Phys. Rev. B 93, 054112 (2016), K. Kanamori et al. Phys. Rev. B 97, 125124 (2018)]. さらに、化合物記述子を利用した無機結晶データベースの結晶構造分類、第一原理計算データベースに基づいた結晶構造探索に取り組み、本研究の化合物記述子が有効であることを実証した[継続中]。

## 2. 新規無機化合物発見のための推薦システム

無機結晶データベースに推薦システム手法を応用することで、合成可能な新規無機化合物を効率的に発見する方法を提案した。具体的には、行列およびテンソル分解によるアルゴリズムにより新規無機化合物の推薦システムを構築する方法を提案した[研究成果(論文発表) 1]。Tucker 分解推薦システムのランキング上位の3元系組成 100 件および 3000 件を用いると、ランダムサンプリングに比べ、1300 倍および 500 倍程度の効率で、合成可能な化合物が存在する組成を発見することができる。4元、5元の組成に対しても同様の結果が得られた。

また、電気陰性度などから導出される化学組成記述子を事前知識として用いる推薦システムの方法を提案した[研究成果(論文発表) 2]。その結果、化学組成記述子を事前知識として用いれば、新規無機化合物の発見を加速できることがわかった。さらに、行列・テンソル分解で用いられる One-hot 表現と化学組成記述子を組み合わせることにより、既知データが少ない対象における推薦システムの高性能化にも取り組んだ。Factorization machines により推薦システムを構築した結果、化学組成記述子を考慮することにより、既知データが少ない対象に対する性能を向上させることができることがわかった。[論文執筆中]

これらの方法により、新規無機化合物の発見が大幅に加速されることが期待される。

このように、本研究項目では、多様な化合物を統一的に記述可能な記述子を導入できた。また、多くの材料科学データに対して本研究の化合物記述子が有効であることを示すことができた。よって、本研究項目の目的は達成できたと考えている。

## 3. 今後の展開

### 研究項目Ⅰ 結晶構造記述子を利用した高精度原子間ポテンシャル構築

研究期間内において、本研究の原子間ポテンシャルは多元系への応用の段階まで来た。しかし、現状の方法では系に含まれる元素の組み合わせに比例した数のモデル係数を推定する必要があるため、多くの元素から構成される系では推定すべき係数の数が膨大になってしまう。今後、このような推定すべき係数の数を抑えながら高精度なポテンシャルを作る方法を導入したいと考えている。また、本研究の原子間ポテンシャルを使って、欠陥を含む構造における原子シミュレーションなどの応用を進めていきたい。

### 研究項目Ⅱ 元素記述子、結晶構造記述子を利用した物性予測モデル構築手法の開発



現在も継続中である無機結晶データベースの結晶構造分類，第一原理計算データベースに基づいた結晶構造探索を進めていきたい。特に，第一原理計算データベースに基づいた結晶構造探索は，採用する特徴量や機械学習手法によって，効率が大幅に変わる。今後，様々な系において効率的に結晶構造探索ができるような汎用的な方法論を模索していきたい。

#### 4. 自己評価

##### 研究目的の達成状況

本研究において設定した目的は，機械学習モデルに適した一般的な記述子を考案すること，および物性予測モデル構築することにより記述子の有効性を実証することであった。研究項目 I では，高精度な MLIP が構築可能な系統的・網羅的な構造記述子およびその生成方法を導入することができた。また，導入した記述子を用い，多くの物質において高精度な原子間ポテンシャルを構築することができた。研究項目 II においては，多様な化合物を統一的に記述可能な記述子を導入でき，多くの材料科学データに対して本研究の化合物記述子が有効であることを示すことができた。よって，本研究項目の目的は達成できたと考えている。

##### 研究の進め方(研究実施体制及び研究費執行状況)

手法開発，第一原理計算によるデータベース構築，プログラミング，機械学習の実行など，本研究の大部分は代表研究者が単独で実行した。一方で，機械学習の専門家との研究会を定期的で開催し，さきがけ「マテリアルズインフォマティクス」の研究者(烏山先生，志賀先生)らと研究協力を実施し，研究代表者だけでは難しい融合研究を実施した。また，機械学習の専門家(京都大学:鹿島先生など)と定期的に交流し，上述の推薦システムなど融合研究としての成果が上がった。

物品としては，主に並列計算機および大容量メモリ搭載計算機を導入した。並列計算機を用いて，大量の第一原理計算を実施した。その結果，本研究で必要となった大きなデータセットを作成することができた。また，導入した大容量メモリ搭載計算機を用いて，大きなデータセットを用いた機械学習を行った。

##### 研究成果の科学技術及び社会・経済への波及効果(今後の見込みを含む)

###### 研究項目 I 結晶構造記述子を利用した高精度原子間ポテンシャル構築

1. 本研究の原子間ポテンシャルは，材料科学や物理，化学の分野において，中心的な役割を果たしてきた従来の原子間ポテンシャルを置き換える革新的なものであり，今後の原子間ポテンシャルの主流になると期待される。
2. 第一原理熱力学計算による熱力学量や物性の算出に基づいた非経験的な材料設計は，材料科学全体において革新的な取り組みであり，本研究の原子間ポテンシャルはそれを達成するための基礎となるものである。本研究の高精度原子間ポテンシャルにより，第一原理熱力学計算の応用範囲が限りなく広がると期待される。
3. 高精度原子間ポテンシャル構築のためのソースコードおよび構築した原子間ポテンシャルを公開することを今後予定しており，本研究の原子間ポテンシャルが広く利用されることが期待される。

###### 研究項目 II 元素記述子，結晶構造記述子を利用した物性予測モデル構築手法の開発



4. 本研究で導入した記述子は、予測する物性に対する経験的・物理的法則を必要としないため、経験的・物理的法則が存在しないような複雑な現象を含んだ物性を簡便に予測することができると思われる。
5. 本研究の物性予測モデルや推薦システムの方法では、物性未知の材料に対して物性を高速に予測することが可能であるため、膨大な数の材料候補の中から材料探索を行うことが可能となる。特に、四元系以上の化合物の数は膨大となる上、実験データの少ない化合物がほとんどである。そのため、四元系以上の材料を対象にした材料探索において、特に威力を発揮すると期待される。
6. 無機化合物の推薦システムにより得られている化学組成に対する合成可能性スコアのリストを公開することを今後予定しており、実験の研究者がリストを簡単に参照できるような環境を整えることができれば、新規無機化合物合成の第一歩として利用されることが期待される。

## 5. 主な研究成果リスト

### (1) 論文(原著論文)発表

- |   |
|---|
| 1. A. Seko, H. Hayashi, H. Kashima and I. Tanaka, "Matrix- and tensor-based recommender systems for the discovery of currently unknown inorganic compounds", Phys. Rev. Materials 2, 013805 (2018).                               |
| 2. A. Seko, H. Hayashi, I. Tanaka, "Compositional descriptor-based recommender system for the materials discovery", J. Chem. Phys. 148, 241719 (2018).  |
| 3. A. Takahashi, A. Seko, I. Tanaka, "Linearized machine-learning interatomic potentials for non-magnetic elemental metals: Limitation of pairwise descriptors and trend of predictive power", J. Chem. Phys. 148, 234106 (2018). |
| 4. A. Seko, H. Hayashi, K. Nakayama, A. Takahashi, I. Tanaka, "Representation of compounds for machine-learning prediction of physical properties," Phys. Rev. B 95, 144110 (2017).   |
| 5. A. Takahashi, A. Seko and I. Tanaka, "Conceptual and practical bases for the high accuracy of machine learning interatomic potentials: Application to elemental titanium", Phys. Rev. Materials 1, 063801 (2017).              |

### (2) 特許出願

研究期間累積件数: 0件(公開前の出願件名については件数のみ記載)

### (3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

#### 主要な学会発表

1. 招待講演, A. Seko, Linearized machine learning interatomic potentials for metals and recommender system for the discovery of unknown materials, Total Energy and Force Methods Workshop 2018 (Cambridge University, Cambridge), 2018/1/9.
2. 招待講演, A. Seko, Applications of machine learning to materials data, IPAM workshop: Optimization and Optimal Control for Complex Energy and Property Landscapes (UCLA, Los Angeles), 2017/10/4.

3. 招待講演, A. Seko, Applications of Machine Learning Techniques to First-principles Data, ENGE 2016 (Korea, Jeju), 2016/11/7.
4. 招待講演, A. Seko, First-principles interatomic potentials via compressed sensing, TMS 2016, 2016/2/18.
5. 招待講演, 世古敦人, 材料科学データにおける機械学習の応用, 日本物理学会共催シンポジウム 2018/3/25.

#### 受賞

1. 世古敦人, 本多記念奨励賞, 第一原理計算に基づいた熱力学計算手法および材料設計手法の開発と応用, 2016/5/27.

#### 著作物

1. A. Seko, A. Togo and I. Tanaka, Descriptors for Machine Learning of Materials Data, in Nanoinformatics (Springer, 2018).
2. A. Seko, K. Toyoura, S. Muto, T. Mizoguchi and S. Broderick, Progress in nanoinformatics and informational materials science, MRS Bulletin 43, 690-695 (2018).
3. I. Tanaka and A. Seko, Toward Materials Discovery with First-Principles Datasets and Learning Methods, Chap. 9 in Information Science for Materials Discovery and Design (Springer, 2016).
4. 世古敦人, 田中功, 第一原理計算と機械学習を用いた材料物性予測, 固体物理 52 (11), 733-741 (2017).
5. 田中功, 松永克志, 大場史康, 世古敦人, 材料電子論入門—第一原理計算の材料科学への応用 (2017).