

機械学習ポテンシャルリポジトリの構築とその応用

Development of Machine Learning Potential Repository and Its Applications

Key-words : Machine learning, Interatomic potential, Crystal structure search, Molecular dynamics

世古 敦人・若井 颯音

Atsuto SEKO and Hayato WAKAI (Kyoto University)

1. はじめに

第一原理計算は、量子力学に基づき、元素と結晶構造により指定される系の電子状態やエネルギーの高精度計算が可能であるため、基礎的な物性、電子・原子レベルでの現象の解釈を得る手段として、材料研究においては不可欠なものとなっている。また、近年の計算機や計算技術の進歩により、第一原理計算の多重実行に基づいた材料スクリーニング、結晶構造探索、熱力学計算、格子振動計算、分子動力学計算など多種多様な応用計算が一般的な手段になってきている。一方で、このような高度な計算は、基礎的な第一原理計算と比べると、計算に必要となるコストは膨大で、実行可能な原子数、時間範囲、自由度は非常に限定的である。

著者は、第一原理的な高度な応用計算を高速化させる手段の一つとして、機械学習ポテンシャルの構築を行ってきた。機械学習ポテンシャルとは、多数の結晶構造に対する網羅的な第一原理計算の結果に基づき、柔軟な機械学習エネルギー予測モデルを用いて、原子間相互作用を表現したものである。このような機械学習ポテンシャルを用いることで、第一原理計算に近い精度のエネルギー計算や力計算が、第一原理計算よりも何桁も小さい計算コストにより、実現可能となる。そのため、これまで実行不可能であった探索範囲の広い結晶構造探索や高精度な大規模分子動力学計算、高精度な有限温度物性計算などが可能となるため、機械学習ポテンシャルを構築するための手法開発や機械学習ポテンシャルの応用が、世界中で活発に行われている。本稿では、このような機械学習ポテンシャルの構築方法や応用例について紹介する。

また、著者により網羅的に構築された多項式型の機械学習ポテンシャルは、現在 48 種の単体や 120 種の合金について、WEB にて公開中 (Machine Learning Potential Repository at Kyoto University)^{1),2)} であり、継続的にポテンシャルが追加される予定である。これらの機械学習ポテンシャルは、世界中で広く用いられている分子動力学計算ソフトウェア LAMMPS にて利用可能な形式であり、LAMMPS で用いるためのパッケージも併せて公開している²⁾。また、それぞれのポテンシャルにより予測された多様な結晶構造に対する凝集エネルギー、エネルギー体積曲線、フォノン状態密度、弾性定数、格子定数などが併せて閲覧できるようになっており、利用可能なポテンシャルの中から、目的に応じて適切なポテンシャルを選択し、利用することができる。

2. 機械学習ポテンシャルリポジトリの構築

機械学習ポテンシャルでは、隣接原子分布を表現する「構造特徴量」を定義し、結晶構造のポテンシャルエネルギーと構造特徴量の関係をガウス過程モデル³⁾、ニューラルネットワークモデル⁴⁾、線形モデル^{5)~8)}などの機械学習モデルにより記述する。多数の構造特徴量と柔軟な機械学習モデルを用いるため、第一原理計算に近い精度での予測が、経験的ポテンシャルに近い計算コストにより可能となる。本稿で紹介する機械学習ポテンシャルでは、球面調和関数に基づいた多項式回転不変量を群論的手法により系統的に生成し、構造特徴量として用いる^{1),8),9)}。また、ポテンシャルエネルギーと構造特徴量の関係を多項式により記述する。そのようなポテンシャルモデルに対し、それぞれの系について、ICSD に収録されているプロトタイプ構造から生成された 1 万から 3 万程度の構造に対して第一原理計算を行い、得られたエネルギーや原子に働く力を学習データとして、線形リッジ回帰によりポテンシャルエネルギーモデルに含まれる係数を推定する。

それぞれの系において最適なポテンシャルを構築するため、カットオフ半径、構造特徴量の数、多項式回転不変量の次数などを変化させたさまざまなポテンシャルモデルを用いてポテンシャルを構築した。これらの予測精度と計算コストはトレードオフの関係にあり、最適なポテンシャルを探すことは、多目的最適化である。よって、最適なポテンシャルはパレート最適なポテンシャルの集合となる。図 1 に、合計 20 種の単体、二元系合金におけるさまざまなポテンシャルモデルの予測誤差と計算コストの関係を示す。パレート最適なポテンシャルと比べ、予測精度と計算コストの両方が

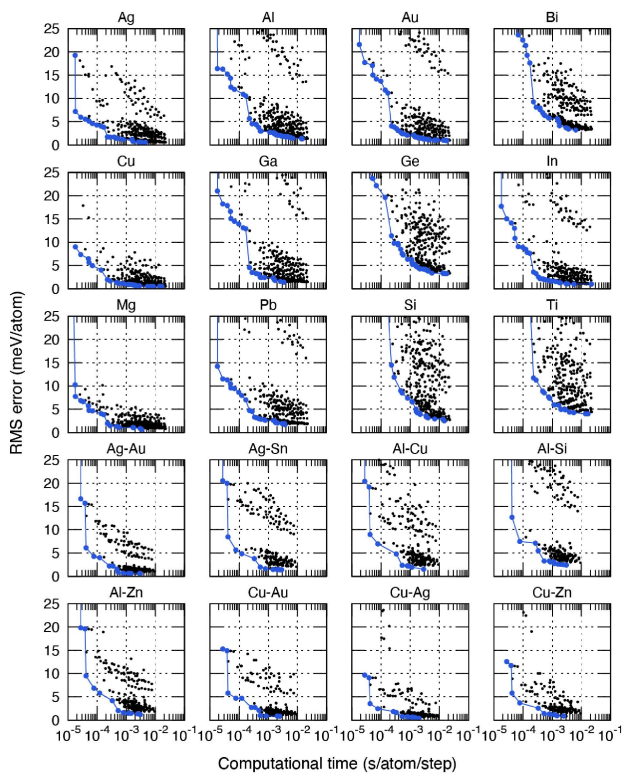


図1 単体および二元系合金における機械学習ポテンシャルの予測誤差と計算コストの関係¹⁾。横軸は、エネルギーと原子に働く力の計算に必要な1原子1ステップあたりの時間。折れ線はパレート最適なポテンシャル。A. Seko, J. Appl. Phys., 133, 011101 (2023); licensed under a Creative Commons Attribution (CC BY) license.

向上するポテンシャルは存在しない。また、図1より、すべての系において、予測誤差が非常に小さいポテンシャルが複数存在することがわかる。これらのポテンシャルは、同じ方法により機械的に構築されたものであるが、柔軟な機械学習モデルおよびその候補モデルの多様性により、どのような系においても高精度なポテンシャルが構築されている。また、これらのポテンシャルのエネルギーや力の計算に必要な時間は、1ステップ、1原子あたり、0.1から10ミリ秒程度であり、広く利用されている経験的ポテンシャルと比べると、計算コストは10倍から1000倍程度となるものの、高精度なエネルギー計算などが可能となる。

このような方法で構築された機械学習ポテンシャルは、多様な結晶構造に対して高精度である。例えば、図2(a)は、Ti-Al系において、さまざまな結晶構造の形成エネルギーおよびその安定構造を機械学習ポテンシャルにより予測したものである。形成エネルギーおよび安定構造の高精度予測には、多数の構造に対する相対的なエネルギーの大小関係をすべて高精度に予測しないとイケないため、ある特定の構造のエネルギー予測と比べてはるかに難易度が高い。図2(a)に示す

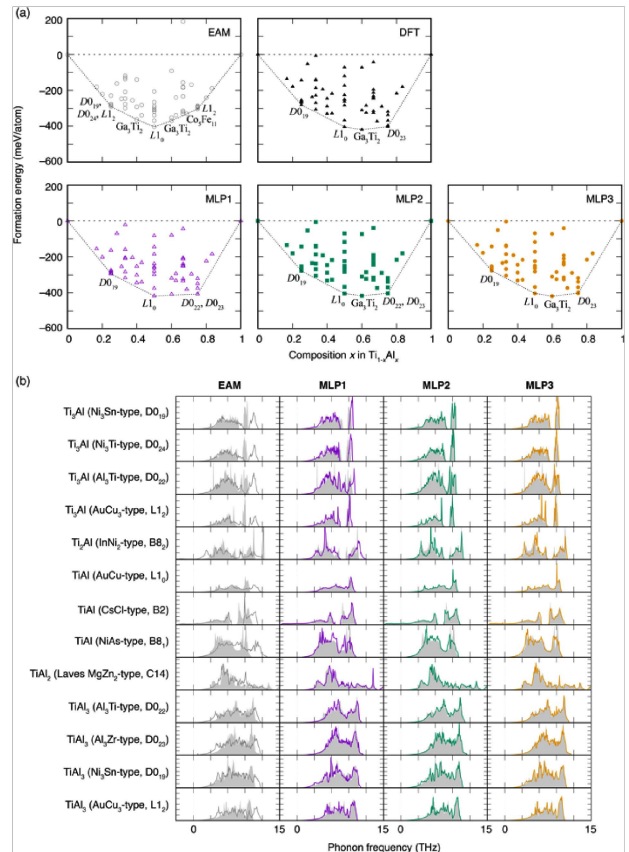


図2 (a) Ti-Al系におけるさまざまな結晶構造の形成エネルギーと安定構造の比較。凸包上の結晶構造が安定構造を示す⁹⁾。(DFT:第一原理計算, MLP:機械学習ポテンシャル, EAM: EAMポテンシャル) (b) 三種類の機械学習ポテンシャルおよびEAMポテンシャルにより予測されたTi-Al系規則構造のフォノン状態密度⁹⁾。(グレーの領域は、第一原理計算による状態密度。)

ように、経験的ポテンシャルの一つであるEAMポテンシャルの安定構造は、第一原理計算の安定構造と比べ、大きく異なることがわかる。これは、EAMポテンシャルのエネルギー予測誤差が大きいことを反映している。一方で、機械学習ポテンシャルは、多数の結晶構造に対する予測誤差が非常に小さいため、第一原理計算と同じ安定構造を予測することができている。また、その他の準安定構造に対する予測も高精度である。

さまざまな原子の配位環境に対するエネルギー予測だけでなく、原子に働く力や関連した物性に関しても、高精度に予測される。例えば、図2(b)は、Ti-Al系において、機械学習ポテンシャルにより規則構造のフォノン状態密度を計算したものである。EAMポテンシャルでは、どの規則構造に対しても、高振動数側の振動数が過大評価されている一方、機械学習ポテンシャルによる計算では、第一原理計算とほぼ同じフォノン状

態密度となる。本稿で紹介した例はごく一部であり、さまざまな結晶構造における原子の運動に関わる物性が高精度に予測される。詳しくは、文献¹⁾を参照されたい。

3. 機械学習ポテンシャルの応用例

このような機械学習ポテンシャルを用いることにより、一般的な第一原理計算では不可能である大規模構造での計算が、第一原理計算と同じような精度で実行可能である。ただし、EAMなどの経験的ポテンシャルと比べると、計算可能な原子数やステップ数は少し減少する。例えば、図3はFCC金属において、 $\langle 110 \rangle$ 対称傾角粒界の構造および粒界エネルギーの予測を行ったものである¹⁰⁾。2112原子までの粒界構造モデルを考え、それぞれの粒界構造モデルについて、さまざまな剛体変位を与えることにより、粒界モデルの大域的な構造最適化を行った。用いた機械学習ポテンシャルは、粒界構造データを一切用いず構築したものであるが、粒界エネルギーの予測値は第一原理計算の値を再現しており、学習データに含まれない欠陥構造に対しても予測能力が高いことを示している。その他にも、10000原子、ステップ数1000000程度の摂動分子動力学計算やPhonon wave packet法計算を通して、Si粒界における格子熱伝導計算を行い、機械学習ポテンシャルが格子欠陥を含む系の構造最適化や物性計算に対して、高精度かつロバストであることを示した¹¹⁾。

また、機械学習ポテンシャルは、不規則状態に対する計算に対しても応用可能である。図4は、Siにおいて、機械学習ポテンシャルを用いた分子動力学計算により液体状態の平均構造を評価したものである¹²⁾。NVTアンサンブルを用いた分子動力学計算によって、液体状態における平均構造の計算を行った。まず常圧下における融点での液体状態の密度に対応する構造を生成した。次に、融点よりも十分に高い温度において、構造を平衡化させ、一つの液体状態を得た。その後、液体状態のスナップショットである構造を出発構造として、複数の温度条件で分子動力学計算を実行した。図4のように、Si、Ge、Beといった異方性の存在する液体に対しては、Tersoffポテンシャルなどの原子間ポテンシャルと比較して、高い予測精度を示している。さらに、液体状態同様、ガラスやアモルファスといった不規則状態に対しても構造や物性の高精度な予測が期待される。

このように、機械学習ポテンシャルの最も分かりやすい応用は、従来は経験的ポテンシャルで計算が行わ

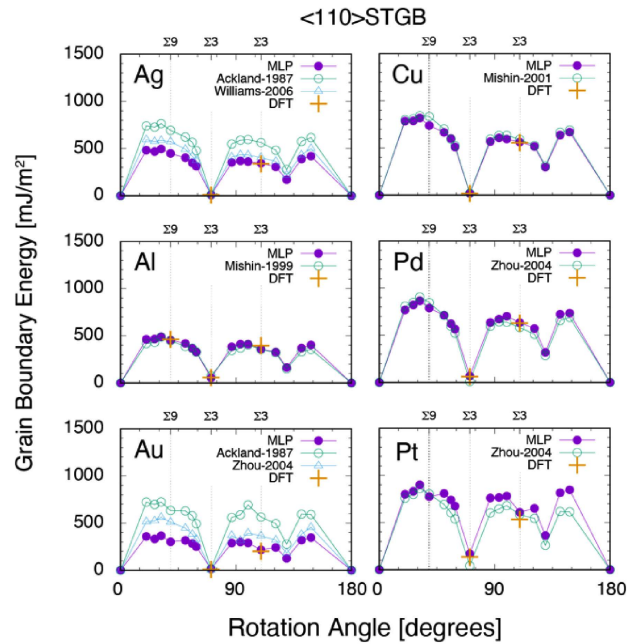


図3 FCC金属における $\langle 110 \rangle$ 対称傾角粒界の過剰エネルギー¹⁰⁾(MLP:機械学習ポテンシャル, DFT:第一原理計算, その他:EAMポテンシャル)

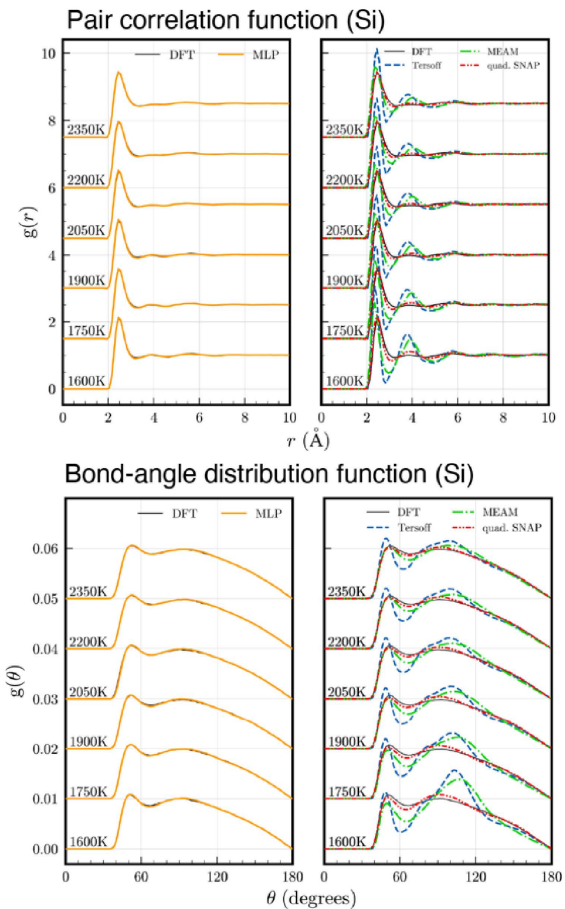


図4 分子動力学法により計算したSiにおける液体状態の構造特徴量¹²⁾。(MLP:機械学習ポテンシャル, DFT:第一原理計算, quadratic SNAP:機械学習ポテンシャルの一種, その他:経験的ポテンシャル)

れていたような大規模構造での構造最適化および分子動力学計算などの高精度化である。一方で、さまざまな結晶構造について、エネルギーや力の高精度な繰り返し計算が必要となる場合にも、機械学習ポテンシャルは非常に有効である。そのような応用例の一つが大域的構造探索である。多くの結晶構造に対して高精度なエネルギー予測が可能であるという特徴を活かし、膨大な数の結晶構造について、機械学習ポテンシャルによるエネルギー計算を実施することにより、化学組成とポテンシャルのみを入力とし、第一原理計算のみでは不可能である探索範囲の広い効率的な大域的構造探索や準安定構造列挙を行うものである。

第一原理計算のみを用いた大域的構造探索は、遺伝的アルゴリズムなどの大域的最適化手法に基づいた構造探索が提案され、最近では、公開されているソースコードの普及により、開発者以外の研究者にも利用され始めている。しかし、安定構造が単純な系を除いては、何らかの構造が出力されるものの、真の安定構造は簡単に見つからないというのが実情である。特に、多元系や高压相など構造が未知である場合や単位胞の原子数が多い構造を対象とする場合、大域的構造探索は非常に難しい。大域的構造探索の必要性や難しさは古くから認識されており、最適化の自由度が原子数に対して指数関数的に増大することが本質的な問題である。

一例として、Al-Cu系における機械学習ポテンシャルを活用した大域的構造探索の例¹³⁾を示す。大域的構造探索では、機械学習ポテンシャルを構築に用いた学習データからかなり離れた構造に対する精度も要求されることが多い。そのため、形成エネルギーの低い構造の探索精度を向上させるべく、機械学習ポテンシャルの改善を実施した。精度改善の手続きとしては、まず、原子数1-8個の範囲で表現できる全組成において、マルチスタート法による大域的構造探索を実行した。多数の初期構造をランダムに生成し、機械学習ポテンシャルにより局所構造最適化を行った。次に、得られた局所安定構造に対して、第一原理計算を実施し、局所安定構造に対する第一原理計算のエネルギーデータなどを学習データに追加することで、機械学習ポテンシャルを再推定した。これにより、形成エネルギーの低い構造の予測精度が向上した。

精度を改善した機械学習ポテンシャルを用いて、原子数1-12、16個の範囲で表現可能な全組成において、マルチスタート法による探索を行った。発見された局所安定構造に対する機械学習ポテンシャルによる形成エネルギーの予測値を図5(a)に示す。エネルギーの

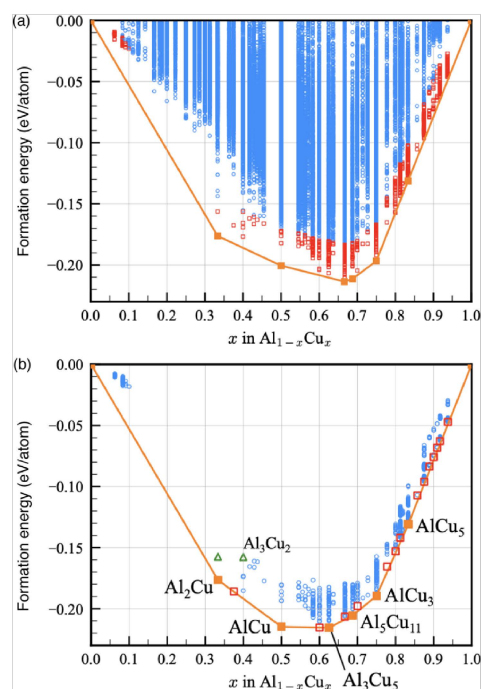


図5 大域的構造探索による Al-Cu 系の安定構造および局所安定構造の形成エネルギー¹³⁾。

計算を実施した構造数は、合計 10^{10} 回程度に到達し、第一原理計算だけでは実施不可能である探索範囲において、大域的構造探索を行ったことになる。一方で、機械学習ポテンシャルには誤差が含まれているため、機械学習ポテンシャルのエネルギー予測値から求まる安定構造は、正しくない可能性がある。よって、発見された局所安定構造から、凸包との形成エネルギー差が 30 meV/atom 以下の局所安定構造を抽出し、第一原理計算による構造最適化を行うことにより、凸包上の構造である安定構造を求めた。図5(b)は、抽出された構造に対する第一原理計算により求められた安定構造および局所安定構造の形成エネルギーである。凸包上の構造においては、実験により報告されている構造と一致する構造が多く発見された。また、実験により報告されている構造と一致した準安定構造もいくつか存在した。このように、機械学習ポテンシャルを利用することにより、第一原理計算の精度は維持した上で、探索範囲を大幅に拡張した大域的な構造探索および準安定構造列挙を実施することが可能である。本稿では、Al-Cu系での大域的構造探索例を紹介したが、同様の方法により、合金系における構造探索だけでなく、圧力下における構造探索、準安定構造が多い系における構造探索などにも応用可能であり、機械学習ポテンシャルを用いた効率的な大域的構造探索がより一般的な形で利用されることが期待される。

謝 辞 本研究は科研費基盤研究 (B) (22H01756) および新
学術領域研究 (19H05787) の助成を受けたものである。

文 献

- 1) A. Seko, *J. Appl. Phys.*, **133**, 011101 (2023).
- 2) A. Seko, <https://sekocha.github.io>
- 3) A. P. Bartók, M. C. Payne, R. Kondor and G. Csányi, *Phys. Rev. Lett.*, **104**, 136403 (2010).
- 4) J. Behler and M. Parrinello, *Phys. Rev. Lett.*, **98**, 146401 (2007).
- 5) A. Seko, A. Takahashi and I. Tanaka, *Phys. Rev.*, **B 90**, 024101 (2014).
- 6) A. Thompson, L. Swiler, C. Trott, S. Foiles and G. Tucker, *J. Comput. Phys.*, **285**, 316 (2015).
- 7) A. V. Shapeev, *Multiscale Model. Simul.*, **14**, 1153 (2016).
- 8) A. Seko, A. Togo and I. Tanaka, *Phys. Rev.*, **B 99**, 214108 (2019).
- 9) A. Seko, *Phys. Rev.*, **B 102**, 174104 (2020).
- 10) T. Nishiyama, A. Seko and I. Tanaka, *Phys. Rev. Mater.*, **4**, 123607 (2020).
- 11) S. Fujii and A. Seko, *Comput. Mater. Sci.*, **204**, 111137

(2022).

- 12) H. Wakai, A. Seko, H. Izuta, T. Nishiyama and I. Tanaka, in preparation.
- 13) H. Wakai, A. Seko and I. Tanaka, *J. Ceram. Soc. Jpn.*, **131**, 762 (2023).

筆 者 紹 介



世古 敦人 (せこ あつと)

2007年京都大学大学院工学研究科博士課程修了。
京都大学次世代開拓研究ユニット助教、京都大学
大学院工学研究科助教を経て、2015年10月より
京都大学大学院工学研究科准教授。専門分野：計
算材料科学。第一原理計算、統計熱力学、機械学
習手法を用いた第一原理熱力学計算手法の開発、
マテリアルズ・インフォマティクス手法の開発に
従事。

[連絡先] 〒606-8501 京都市左京区吉田本町
京都大学大学院工学研究科材料工学専攻
E-mail : seko@cms.mtl.kyoto-u.ac.jp