

機械学習による結晶構造安定性の高速予測

Fast Prediction of Crystal Structure Stability by Machine Learning

*(M1)稲田 祐樹¹, 桂 ゆかり^{1,2,3}, 熊谷 将也^{3,4}, 木村 薫¹¹ 東大院新領域 ² 物材機構 ³ 理研 ⁴ さくらインターネット(株)¹ The University of Tokyo. ² National Institute for Materials Science. ³ RIKEN. ⁴ Sakura Internet Inc.

E-mail: inada@phys.mm.t.u-tokyo.ac.jp

1. 緒言 新規材料の探索では、目的の物質の結晶構造が安定であるかを判別し、その性質を予測することが求められる。多くの元素の組み合わせを効率よく探索するためには、結晶構造の安定性の高速な予測が必要である。

結晶構造の安定性は、その結晶構造の生成エネルギー E_{str} と、その組成での最安定構造のエネルギー E_{min} の差 ΔE により評価される。第一原理計算による結晶構造予測では、多くの競合仮想結晶構造のエネルギー計算を行う必要があり、計算コストが大きい。

第一原理計算の計算結果を教師データとする機械学習によって、その安定性を高速に予測できることが期待される。しかし現在は、競合構造間の安定性の比較に十分な生成エネルギー E の予測精度は得られていない。また、 ΔE を直接教師データとして学習した場合にも、その予測は困難であった。この原因として、入力記述子と、 E または ΔE の機械学習プロセスに課題があると考えられる。

機械学習には、元素の性質を数値として表現した元素記述子が必要である。これまでは単体元素の物性値が広く用いられてきたが、原子半径や電気陰性度など少数の本質的な記述子を除けば、結晶構造中での原子の挙動と直接関係しない記述子が多かった。一方、より本質的な、元素記述子の作成手法として Atom2vec[1]という手法が考案されている。これは、結晶構造データベース中の数万化合物の組成比の頻度行列を特異値分解することで、各元素の類似性を数値化した記述子を自動的に作成する手法である。ここから、組成以外の構造情報の頻度行列にも同様の特異値分解を施せば、より多くの情報を持った元素記述子が作成できると考えた。そこで本研究では、化学組成に加え、結晶構造中での配位多面体に着目して独自の記述子を作成した。

E_{str} 及び ΔE の予測には機械学習プログラムへの結晶構造情報の入力が必要である。これは長らく困難であったが、最近、Xieらによって結晶構造を原子同士の結合関係によるグラフ構造(Crystal Graph, CG)で表現する手法が考案された。これによりニューラルネットワーク(NN)に結晶構造を入力することが可能となった[2]。本研究ではCGを入力としたNNにより E_{str} 、組成を入力としたNNにより E_{min} を予測し、その差から ΔE の予測を試みた。

2. 方法 教師データとして使用するため、Materials Project データベース[3]から、128,175化合物の化学組成と原子座標、第一原理計算によるエネルギーのデータを取得した。

そのうち、安定($\Delta E=0$)の化合物のデータは35,746件であった。ToposPro[4]を用いて安定化合物の結晶構造中の各原子の配位原子及び配位多面体のトポロジーのデータを取得した。

化合物中の各元素について、化学組成、配位元素、配位多面体のトポロジーから、特異値分解によって三種類の元素記述子を作成した。これらの記述子をそれぞれ記述子A、B、Cとする。なお、記述子AはAtom2vecの再現であり、B、Cが独自の記述子である。それぞれの元素記述子について、特異値が大きい順に40個ずつを使用した。また、化学組成の特徴を元素記述子で表現するため、各記述子の組成比に対応する重みをかけた和を計算した。

結晶構造の安定性の予測のために、二種類の機械学習モデルを作成した。一つ目は、化学組成記述子を入力としたNN(NN-cmp)で、 E_{min} の予測に用いた。二つ目は、CGを入力としたNN(NN-str)で、 E_{str} の予測に用いた。

NN-cmpを使用して、安定化合物の生成エネルギー(E_{min})の機械学習を行った。NN-cmpによる機械学習プログラムを50回実行し、その予測精度(平均絶対誤差、MAE)の平均によって、記述子による予測精度の差を比較した。比較として、従来型の元素記述子(pymatgen[5])による予測も行った。

また、NN-strとNN-cmpによる予測結果の差を用いて ΔE の予測を行った。

3. 結果・考察

E_{min} 予測精度の記述子依存性を表1に示すが、今回作成した三種の記述子を組み合わせるのが最も高い予測精度を示した。また、 E_{str} と E_{min} の予測の差から ΔE を予測した結果、MAEが0.082eVという高い予測精度を実現できた。この機械学習モデルでは一分間に 10^4 程度の結晶構造の安定性を推定することができた。

4. 結論 本機械学習手法により、結晶構造からその安定性を予測することが可能となった。これにより、安定に近い結晶構造を高速で選別することが可能となり、新規材料探索の効率化に繋がる。今後は、エネルギー以外の第一原理計算データの利用や、元素記述子の改善などにより、結晶構造の安定性をより精度良く判別できる機械学習プログラムの開発を目指す。

5. 参考文献 [1] Zhou, Q. et al., *Proc. Natl. Acad. Sci.* **115**, E6411–E6417 (2018). [2] Xie, T. et al., *Phys. Rev. Lett.* **120**, 145301 (2018). [3] A. Jain, et al., *APL Materials*, **1**(1), 011002 (2018). [4] Blatov V. A. et al., *Cryst. Growth Des.*, **14**, 3576–3586 (2014). [5] S. P. Ong et al., *Comput. Mater. Sci.*, **68**, 314–319 (2013).

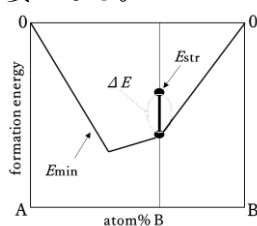


Fig. 1. evaluate crystal structures stability from ΔE

Table. 1. MAE of E_{min} prediction

元素記述子	MAE(E_{min}) (eV/atom)
pymatgen	0.109
記述子 A	0.053
記述子 B	0.079
記述子 C	0.065
A+B+C	0.033