

機械学習による有機材料の結晶構造予測

Prediction of Crystal Structures of Organic Materials by Machine Learning

岩手大院総合, °(M2) 高橋昌也, 吉本則之

Iwate Univ., °(M2) Masaya Takahashi, Noriyuki Yoshimoto

E-mail: g0320091@iwate-u.ac.jp

はじめに：結晶構造は材料の物性を決定づける基本的要素であり、分子構造から結晶構造を予測することは機能性材料の開発を効率化する本質的な手法となる。結晶構造の予測には第一原理計算や分子力学による計算が用いられてきたが、これらの手法は計算過程が複雑であることから多くの時間を要し、また、材料から構造を推定する順方向の推定は可能であっても、目的の構造を持つ材料を探索する逆方向の推定は困難であるという問題があった。そこで本研究では機械学習を用いて有機材料の分子構造から結晶構造を規定する空間群の予測を行う手法の開発を行う。これにより、従来よりも短時間で効率的かつ逆方向の推定が可能な空間群予測方法の確立を試みる。

実験: Saulius Gražulis らが作成したデータセット^[1]から 54,008

個の有機分子を選択し、化学構造データと空間群データを用いてデータセットの作成を行った。化学構造データは

SMILES として取得し、PaDEL-Descriptor^[2]を用いて 9 種類の Finger Print に変換を行った。機械学習は Python の scikit-learn から 4 種類のアルゴリズムを利用し、機械学習のモデルの精度 (Accuracy) とモデルの評価値 (F1-score) の導出を行った。

出力結果である空間群はブラベ格子を BV、主軸に平行な対象要素を T1、主軸以外の対象要素を T2、さらに異なる主軸周りの対象要素を T3 という名称で分割し、それぞれ別々の出力としてデータセットを作成し、機械学習モデルの作成を行った。

結果： Fig.1 にデータセットの割合、Fig.2 に機械学習モデルの最大の Accuracy を示す。また、Fig.3 にそれぞれの機械学習モデルにおける上位 3 位までの F1-score の結果を示す。この結果から Fig.1 に示すように不均衡データセットにおいて、化学構造データと空間群データの関係を学習し、Fig.2 に示すようなモデル精度で、Fig.3 に示すような F1-score で機械学習

による分類が可能になった。これにより、機械学習を用いた有機材料の空間群予測では、少数データの空間群予測は困難であるが、多数データの空間群予測では予測可能であると示唆される。

参考文献

[1] Gražulis, S. et al., *Materials Informatics*, 1–39, John Wiley & Sons, Ltd (2019).

[2] Yap, C. W., *Journal of Computational Chemistry*, **32**, 1466–1474 (2011).

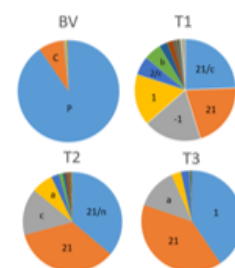


Fig.1 Percentage of data set

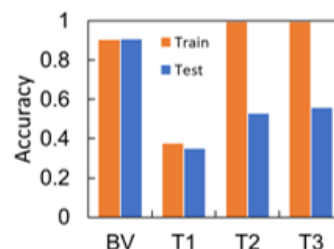


Fig.2 Maximum accuracy of machine learning

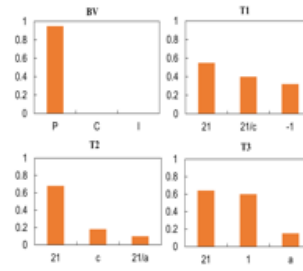


Fig.3 F1-score for each model