



深層学習および転移学習を活用した 小規模データによる化合物探索技術

(材料技術研究部)

1 技術の概要

○背景・目的

異種材料間の馴染みやすさを改善する手段のひとつに、対象物質に親和性の高い官能基を付与する表面修飾がある。各種シミュレーションにより修飾物質と対象物質の配位エネルギーを算出し、それによって親和性を評価することは可能であるが、低分子化合物は膨大に存在するため、最適な物質を探索することは困難である。

そこで本研究では、機械学習の活用により、化合物探索の高速化を試みた。具体的には、化学構造式を入力すれば特定材料との配位エネルギーが出力される予測器の作成を行った。

○研究方法

低分子化合物のデータベースQM9からランダムに抽出した32化合物について、分子動力学(MD)シミュレーションにより対象物質との配位エネルギーを計算した。データ件数が少なく、そのまま予測器を作成することは不可能なため、転移学習を適用した。

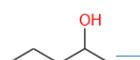
転移元として、深層学習による回帰モデルを作成した。QM9に記載の化合物からランダムの抽出した5,000件の化合物について、化学構造式を説明変数、熱容量を目的変数とし、決定係数は0.964となった。

○結果と考察

パラメータの微調整(ファインチューニング)、学習率やエポック数の設定により、転移先の決定係数は0.659となった。データベースを活用した転移学習により、小規模データからでも、一定水準の予測器を作成可能であることがわかった。

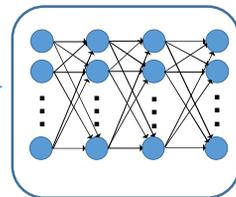
転移元

化学構造式



x5,000件

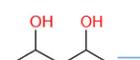
学習済みの
深層学習モデル



化合物の
熱容量
(目的変数)

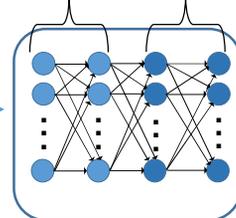
転移先

化学構造式



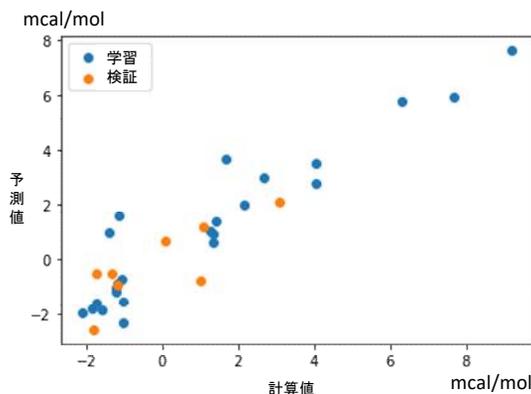
x32件

固定 微調整



配位
エネルギー
(目的変数)

転移学習(ファインチューニング)のイメージ



配位エネルギーの予測

横軸: MDシミュレーションによる計算値
縦軸: 予測器が出力した値
青色: 学習データ(24件)
黄色: 検証データ(8件)

2 このようなお困りごとを解決できます

- ・特定の材料になじみのよい化合物を高速・網羅的に探索したい
- ・化学構造を入力データとして機械学習を行いたい
- ・小規模データを用いて機械学習を行いたい