



機械学習を活用した 高速かつ網羅的な化合物探索技術

(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。量子化学計算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133,886件からなるデータベースQM9より化合物データを取得し、構造の類似性によってクラスタリングを行った。

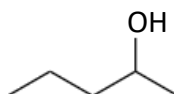
各クラスターからランダムに抽出した化合物に対して候補物質との配位エネルギーを計算したのち、化学構造を多次元のベクトルに変換したフィンガープリントを説明変数、配位エネルギーを目的変数とする教師あり学習を実施した。

○結果と考察

配位エネルギー計算の件数と精度に問題があったため、目的変数として使用することは断念した。

化学構造を説明変数とした教師あり学習の枠組み自体を作成することはできた。例えばHOMOエネルギー(電子供与性の指標)を目的変数とした場合、アルゴリズムおよびパラメータの最適化によって、決定係数0.927の予測器となった。

化学構造式



文字列記法

CCCC(C)O

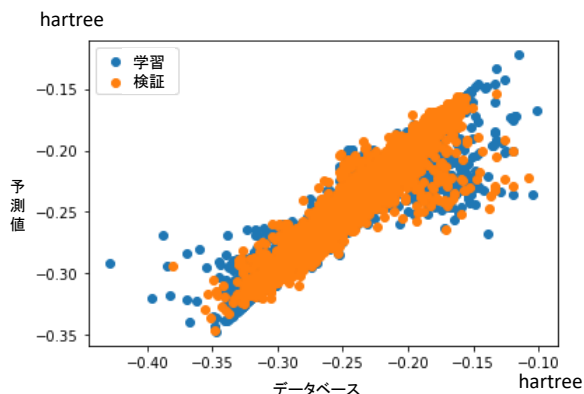
特定物質との配位エネルギー
(目的変数)

学習モデル

フィンガープリント
(説明変数)

0	1	0	0	...
...	0	0	0	1

教師あり学習のイメージ



HOMOエネルギー計算の例

横軸: データベース中の値
縦軸: 予測器が出力した値
青色: 学習データ(100,415件)
黄色: 検証データ(33,471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