3P08

配座データを用いたニューラルネットワークによる分子物性予測

○立花 尚登[1]，五十幡 康弘[1,2]，後藤 仁志[1,2]
1. 豊橋技科大院工（〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘）
2. 豊橋技科大 IMC（〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘）

【緒言】

分子物性を予測するための機械学習モデルを構築する際，通常，一つの分子につき一つの分子グラフや立体構造を用いるが，その配座異性体のアンサンブルは考慮されないことが多い．柔軟な分子は配座異性体を始めとする複数の立体配置で存在しているため，これらを考慮したデータを元に機械学習を行うことで，より多様な物性を高精度に再現できると考えられる．本発表ではそれを検証するため，配座探索を行った結果得られる最安定構造のみを含むデータセット及び，配座異性体のアンサンブルからなるデータセットについて機械学習モデルを学習させ，その予測精度を比較する．

【方法】

公開データベース PubChem[1] より構成原子が B, C, N, O, F, P, S, Cl, Br, I であり分子量が 100 以下となる分子を収集し，CONFLEX[2] により配座を創出，ωB97X-D/6-311+G(d,p) レベルの密度汎関数理論（DFT）計算で構造最適化と基準振動数解析を行う．機械学習モデルへ入力するために，これを 3 次元記述子によりベクトルとして符号化する．本研究では，電子回折の研究で使用される方程式を単純化した式で表現される 3D-MoRSE Descriptors[3]を用いる．計算された記述子ベクトルを入力に，全結合層からなるニューラルネットワークを用いて全エネルギー，HOMO，LUMO，HOMO-LUMO ギャップなどを目標値として学習し，Test セット内の分子について予測を行う．

【結果】

ここでは，HOMO エネルギーについて行った予測について示す(図 1)．配座探索を行った結果得られた最安定構造のみを学習に用いた場合を(1), エネルギーの低い構造から最大 10 個の配座異性体を学習に用いた場合を (2) に示している．Train, Test のどちらの場合においても，配座異性体を学習に用いた (2) の方が DFT による HOMO エネルギーを再現できていることがわかる．決定係数 R2, 絶対平均誤差 MAE, 平方平均自乗誤差 RMSE のいずれによる評価においても高いパフォーマンスを示しており，配座異性体を学習データに加えることで機械学習の精度が向上していることが分かる．現在このような解析を更に進め，他の記述子や物性に対する予測に関して，配座探索はどのような効果をもたらすのか調査している．

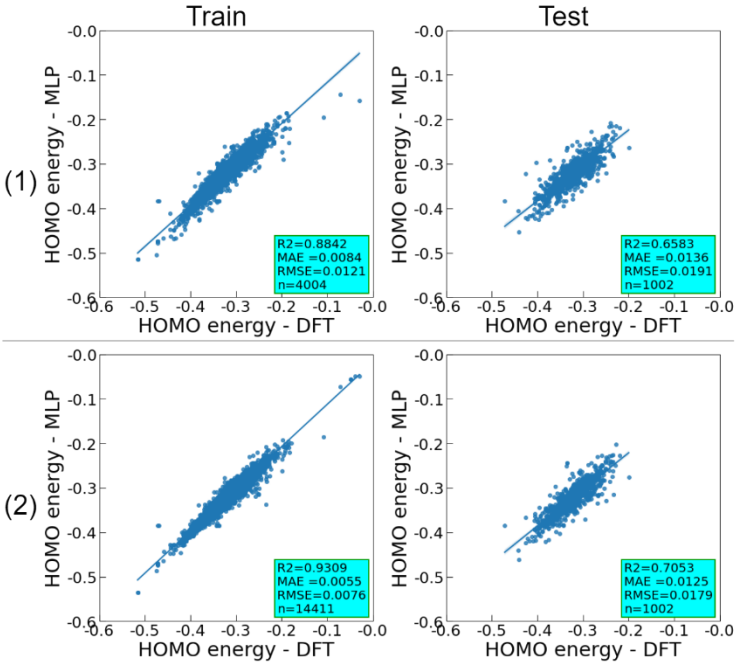

図 1. HOMO エネルギーの学習データ（DFT）と予測結果（MLP）の比較（hartree 単位）．(1) 最安定構造のみの学習データ，(2) 配座異性体を加えた学習データ．MAE, RMSE の単位は Hartree

本研究の一部は豊橋技術科学大学 HPC クラスターおよび東京工業大学 TSUBAME3.0 を利用して実施した。

【参考文献】

[1] Kim, S. et al. Nucleic Acids Res. 2019, 47(D1), D1102-D1109.
[2] Goto, H.; Obata, S.; Nakayama, N; Ohta, K., CONFLEX9, Conflex Corp. 2021.
[3] Devinyak, O.; Havrylyuk, D.; Lesyk, R. J. Mol. Graph. Model. 2014, 54, 194-203.