

信頼される AI の基盤技術
2021 年度採択研究者

2021 年度 年次報告書

横川 大輔

東京大学大学院総合文化研究科
准教授

化学的知見を生かした転送性の高い特徴量の抽出と利用

§ 1. 研究成果の概要

本研究課題は「化学的知見を生かした転送性の高い特徴量の抽出と利用」であり、得られた結果に対する説明性の解明に重きを置いている。

2021 年度は化学的知見を生かした前処理層の提案とその数値的検証を行った。Graph Convolution Networks (GCN)における畳み込みのステップでは、隣接行列を利用することで、各原子の情報を分子全体に伝搬させる。例えば、Wu らが提案した Simple Graph Convolution (SGC)での畳み込みステップでは、規格化された隣接行列の多項式を用いており、シンプルなモデルでありながら、これまでに提案されてきた GCN に基づいた手法と遜色ない結果を示すことが報告された。しかしその一方で、畳み込みの層を増やすと平滑化が進行しすぎてしまい、結果の改善がみられない問題も指摘されていた。

本研究では、規格化された Wiberg 結合次数を利用した畳み込みを提案した。この畳み込みは前述の SGC と同じ結合情報を利用した前処理ではあるが、出力結果は両者で大きく異なる。SGC の場合、層の数を増やすと畳み込まれた特徴量が、ある一定の値に収束するのに対し、本モデルでは層の数が分子内で最も遠く離れた原子間の結合の数よりも大きくなると出力は 0 になる。このように、畳み込みで得られたベクトルの張ることができる空間を制限することで、畳み込み層の数に対して安定なモデルを構築できると考えた。

構築したモデルの妥当性を調べるために、酸解離定数(pKa)を対象にした教師あり学習を行った。本手法は、既存の pKa 予測プログラムと同程度の誤差で pKa を予測できることがわかった。さらに、Integrated Gradients により、予測に際して重要な寄与を示した特徴量を調べたところ、pKa が小さい時は着目する水素サイトの電荷、分子ポテンシャルが重要であることが確認された。