

機械学習によるセリウム錯体の励起エネルギーの予測

慶大院理工 ○荻谷 晃輔, 稲垣 泰一, 畑中 美穂

【序】ランタノイド (Ln) 錯体の一つであるセリウム (Ce) 錯体は光還元触媒としての報告例^[1]を皮切りに次世代の光機能性材料として注目されている。Ce³⁺の電子配置は[Kr]4d¹⁰4f¹5s²5p⁶であり、基底状態と擬縮退した6個の4f¹励起状態よりも高エネルギー側にある第7励起状態が重要な役割を果たすことが多い。ここで4f、5d軌道の性質に着目すると、Ce³⁺の5d軌道は配位子場の影響を強く受けるが、4f軌道は閉殻5s²5p⁶電子に遮蔽されるため、配位子場の影響を受けにくいという特徴がある。そのためCe³⁺錯体のf-d遷移の励起エネルギーは5d軌道のエネルギーから予測できると考えた。そこで本研究では、Ce³⁺錯体の基底状態の情報から第7励起エネルギーを予測する機械学習モデルを作成し、励起エネルギーに寄与する因子の抽出を目指した。

【計算】ケンブリッジ結晶構造データベース^[2]より684種のLn (= Eu, Tb)単核錯体を集め、LnをCeに置換した。得られた錯体構造に対して、PM6-D3/Sparkle法^[3]による構造最適化を行い、一点計算にはDFT法、TD-DFT法を用いた(汎関数はωB97XD)。Ce³⁺には相対論的有効内殻ポテンシャル(RECP)とそれに対応する基底関数を適用し、そのほかの原子にはcc-pVDZ基底関数を使用した。特に、励起状態計算では”small core RECP”を用い、基底状態計算では”large core RECP”を用いた^[4,5]。機械学習モデルには、線形回帰やLinearTree回帰^[6]を用いた。多重共線性の問題を回避するために、互いの相関係数が小さい13個の説明変数を用いた。なお、LinearTreeモデルの分岐には我々が定義した48個の部分構造の個数を用いた。

【結果】最初に最も単純な線形重回帰モデルを構築したところ、検証データに対する相関係数 R^2 が0.77となり、比較的良好な予測精度を示した。一方、得られたモデルの回帰係数を見ると、予想に反して5d軌道エネルギーでなく、HOMO-LUMOギャップの方が励起エネルギーに高く寄与することが分かった。そこで、基底状態と第7励起状態におけるCeのpopulationを調べたところ、これら遷移にはf-d遷移以外に、f-f遷移や配位子内励起、電荷移動型励起が混在していると分かった。このことから、励起の種類に応じて重要な因子が異なると予想できる。そこで、データの分岐と回帰を行うLinearTree法を用いた。LinearTree法は、回帰木やランダムフォレスト(RF)^[7]などの分岐先でデータの平均値を返すモデルとは異なり、分岐先で回帰モデルを作るため、外挿性が良いと考えられる。実際にモデルの精度を比較したところ、RFモデルの場合、訓練データとテストデータに対する相関係数 R^2 が0.98、0.77で過学習していたのに対し、LinearTreeモデルの場合、スコアが0.82、0.81と堅牢なモデルができた。加えて、LinearTreeモデルは単一の木構造からなるモデルであるため、複数の木構造からなるアンサンブル学習による精度向上が期待できる。発表ではLinearTreeモデルとアンサンブル学習を組み合わせた結果についても議論する。

【参考文献】

- [1] Haolin Yin, et. al., *J. Am. Chem. Soc.*, **2015**, 137.
- [2] C. R. Groom, et. al., *Acta Cryst.* **2016**, B72, 171.
- [3] R.O. Freire, A.M. Simas, *J. Chem. Theory Comput.* **2010**, 6, 2019.
- [4] M. Dolg, H. Stoll, H. Preuss, *J. Chem. Phys.* **1989**, 90, 1730.
- [5] M. Dolg, H. Stoll, H. Preuss, *Theor. Chim. Acta* **1993**, 85, 441.
- [6] <https://github.com/cerlymarco/linear-tree>.
- [7] Breiman, L. *Mach. Learn.* **2001**, 45.

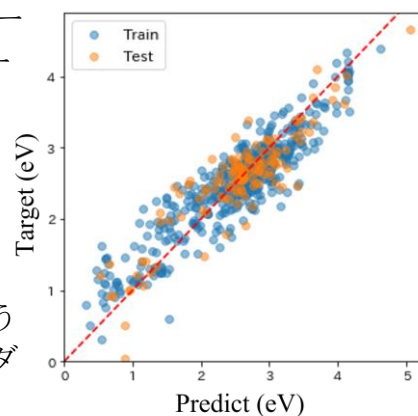


Fig. 1. yy-plot of the LinearTree model