

機械学習を利用した有機半導体・蛍光体分子設計支援ツールの開発

Machine-Learning-Assisted Design Tool for Semiconducting and Fluorescent Molecules

山形大 ROEL¹ °松井 弘之¹, 岡田 智悠¹

ROEL, Yamagata Univ.¹ °Hiroyuki Matsui¹, Tomoharu Okada¹

E-mail: h-matsui@yz.yamagata-u.ac.jp

有機半導体や蛍光体などの分子設計において、最高被占軌道 (HOMO) および最低空軌道 (LUMO) エネルギーは電気伝導特性や光学特性等を決定する重要なパラメータである。現在、計算機上で HOMO, LUMO エネルギーを求める方法の主流は密度汎関数理論 (DFT) 計算であるが、構造最適化も含めると計算時間は一般的に数分～1 時間程度を要する。本研究では、機械学習によって分子構造式から 3 次元構造データを介することなく瞬時に HOMO, LUMO エネルギーを予測し、それを構造式エディタと連携させることにより、ヒトが分子を描画する際にリアルタイムで HOMO, LUMO エネルギーの予測値をフィードバックする分子設計支援ツール (Fig. 1) を開発した。

学習データとして、ケンブリッジ結晶構造データベース (CSD) に含まれる有機分子の 3 次元構造を元に DFT 計算 (B3LYP/6-31G*) を行い、約 22 万件の分子構造式および HOMO, LUMO エネルギーの組み合わせを得た。更に、分子構造から部分木構造に基づいて特徴量抽出を行い、Lasso 回帰によって機械学習を行った。例えば、Fig. 1 中の分子の硫黄原子に関する部分木構造は S-CC または S-(C-CCS)(C-CCS) と表すことができる。その後、得られた学習モデルをオープンソースの構造式エディタ JSME と統合することにより、構造式の描画と同時に HOMO, LUMO エネルギーの予測値を表示可能なツールを作成した (Fig. 1)。最後に、18 名のボランティアに本ツールを使用してもらい、その有効性や予測精度について検証した。

ケンブリッジ結晶構造データベースを元に作成した学習データおよびテストデータに対する HOMO 予測の平均絶対誤差 (MAE) はいずれも 0.28 eV であったが、一方でボランティアが新たに設計した 108 種の分子に対する MAE は 0.49 eV であり (Fig. 2)、データベースから生成したテストデータよりもやや低いスコアとなった。これはデータベース上の分子とボランティアが設計した分子の集合にそれぞれ偏りがあるためと考えられる。また、ボランティアに n 型有機半導体に適した LUMO の深い分子を設計してもらったところ、支援ツール使用時の方がより LUMO の深い分子を得られる傾向が見られた。本ツールは近日中に Web で公開予定である。

謝辞：本研究は、JST, CREST, JPMJCR18J2 の支援を受けたものである。

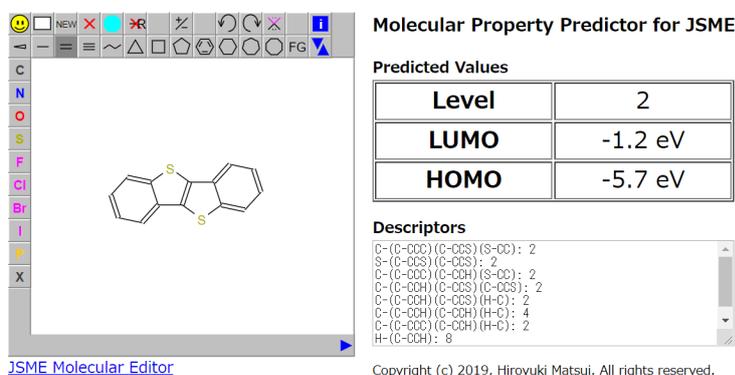


Fig. 1 Screenshot of the machine-learning-assisted design tool for semiconducting and fluorescent molecules

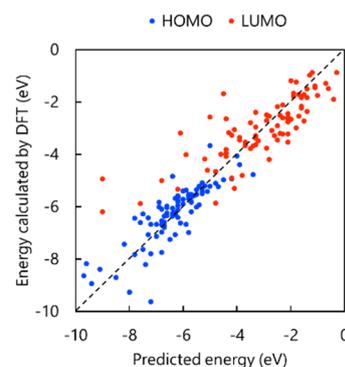


Fig. 2 Machine learning vs DFT for the molecules designed by volunteers