機械学習による有機半導体分子の HOMO, LUMO 予測の高精度化

Improving the Accuracy of HOMO/LUMO Prediction of Organic Molecules by Machine Learning

山形大 ROEL ○(M1)望月 祐樹， 岡田 智悠, 松井 弘之

Research Center for Organic Electronics (ROEL), Yamagata University

○Yuki Mochizuki, Tomoharu Okada, Hiroyuki Matsui

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

[はじめに] 有機デバイスの性能には半導体の HOMO, LUMO 準位が大きく関係する。有機電界効果トランジスタを例に挙げると、p 型半導体の HOMO 準位と電極の仕事関数が近いほど接触抵抗が小さくなり素子性能が向上する。当研究室では既に機械学習を用いて有機半導体の HOMO, LUMO 準位を高速に予測し、分子設計にフィードバックするツール「Yamagata Chemical Canvas」（https://matsui-lab.yz.yamagata-u.ac.jp/softwares.html）を開発・公開しているが、予測精度が低いという課題があった。本研究では、より高精度に有機半導体分子の HOMO, LUMO 準位を予測することを目的とした。

[研究方法] The Cambridge Structural Database からエラーを含まないなどの条件で 315,519 個の分子を抽出した。抽出した分子のうち電子数の小さいものから 108,400 分子を取り出し、Gaussian を用いて構造最適化と DFT 計算を行った。DFT 計算ができなかった分子を除いた 106,676 分子の HOMO, LUMO 準位が決定した。これらの分子から木構造とフラグメントという 2 種類の手法を用いて部分構造をそれぞれ抽出した。木構造は基準となる原子と隣接した原子をまとめて分子の部分構造として表現した。フラグメントは分子を単結合で切断し、分割された構造を分子の部分構造として表現した。HOMO, LUMO 準位を目的変数、2 種類の部分構造の出現回数を説明変数として Lasso 回帰と Ridge 回帰を行い、予測モデルの作成と評価を行った。

[結果・考察] Ridge 回帰による HOMO, LUMO 準位の予測値と DFT 計算値の相関を Fig. 1, Fig 2 にそれぞれ示す。テストデータに対する平均絶対誤差が HOMO は約 0.215 eV、LUMO は約 0.297 eV となった。構造最適化をせず、記述子に木構造のみを採用していた先行研究と比較すると、平均絶対誤差が HOMO は約 24 %、LUMO は約 17 %改善した。改善の理由として、DFT 計算を行う前に原子の 3 次元座標の最適化を行い実験的な誤差を少なくしたこと、機械学習に使用する記述子の種類を 2 種類に増やしたことなどが挙げられる。また、n 型有機半導体に有効な LUMO 準位が深くなるフラグメントを抽出したところ、Fig. 3 のフラグメントが得られた。

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

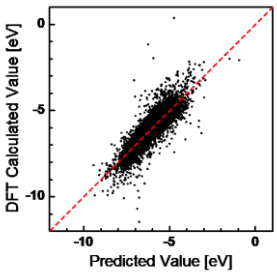

Fig 1. HOMO's scatter plot

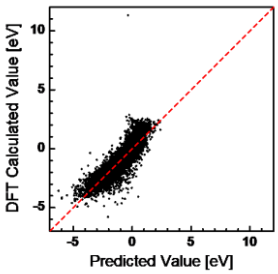

Fig 2. LUMO's scatter plot

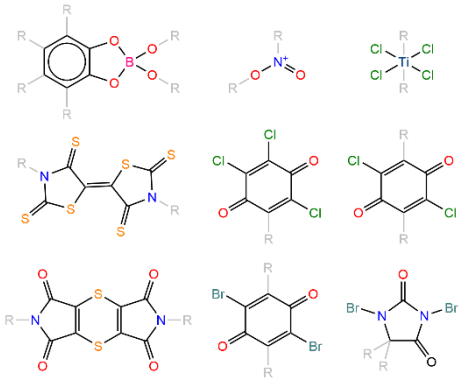

Fig 3. Fragments deepening the LUMO