

解釈可能な機械学習を活用した 化学組成によるバンドギャップ予測

Prediction of band gaps from chemical compositions

using interpretable machine learning

京大工 °(B) 増田 太一, 田辺 克明

Kyoto Univ., °Taichi Masuda, Katsuaki Tanabe

E-mail: tanabe@cheme.kyoto-u.ac.jp

背景と研究概要

バンドギャップは、半導体材料の光学的、電子的特性を決定づける重要なパラメータである。しかし、長年にわたり、計算コストを抑えつつバンドギャップを正確に理論計算することは大きな課題とされてきた。この課題を解決するため、データ駆動型の機械学習手法を用いたバンドギャップ予測に関する研究が進められてきた。しかし、機械学習の予測精度と解釈性の間にはトレードオフの関係が存在するため、高い予測精度を有する複雑なモデルは、予測結果の解釈が困難であることが多い。本研究では、説明可能な人工知能(XAI)の技術を活用して、化学組成からバンドギャップを予測する機械学習の分析をした。

計算手法

本研究では、Permutation Feature Importance(PFI)や、Partial Dependence Plot(PDP)、Individual Conditional Expectation Plot、Accumulated Local Effects Plot、Shapley Additive Explanations といった XAI の技術を活用した。これらの手法を用いて、バンドギャップの実測値^[1]および密度汎関数理論によるバンドギャップの計算値^[2]を予測するサポートベクター回帰(SVR)、勾配ブースティング回帰、およびランダムフォレスト回帰モデルを分析した。

結果と考察

上記の XAI の手法により、バンドギャップ予測モデルにおける重要な物理量や、特徴量として使用された化合物構成元素の特性とバンドギャップの依存関係を明らかにした。具体的には、PFI をバンドギャップ予測モデルに適用することで、共有結合を形成する平均電子数の平均や平均質量密度がバンドギャップ予測モデルにおける重要な特徴量であることを示した。また、PFI によって導出された重要な特徴量はモデルの予測性

能に大きな影響を与えることを示した。

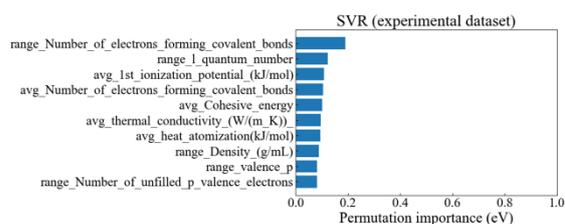


Fig. 1 PFI for SVR with the expt. dataset.

さらに、PDP の結果から、化合物の元素の平均質量密度が増加すると、バンドギャップが減少するという依存関係を明らかにした。特に、分析を行なった3つの機械学習モデルの中で、SVR の PDP が構成元素の特性とバンドギャップの関係を捉えるのに最も適していることがわかった。これらの結果は、材料の電子構造の観点から理論的に解釈することができた。また、PFI に基づいた特徴量選択によって、バンドギャップを予測する機械学習の予測精度が向上した。^[3]

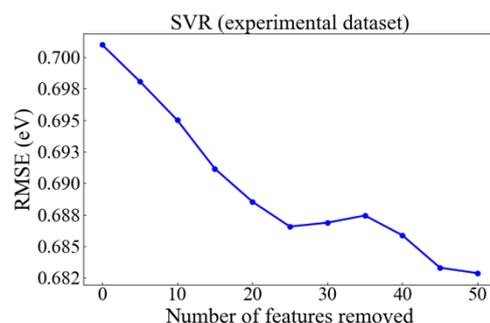


Fig. 2 Relationship between the number of removed features and RMSE for SVR with the expt. dataset, with features dropped from lowest to highest PFI importance.

参考文献

- [1] Y. Zhuo, A. M. Tehrani, and J. Brgoch, *J. Phys. Chem. Lett.* **9**, 1668 (2018).
- [2] A. Jain *et al.*, *APL Mater.* **1**, 011002 (2013).
- [3] T. Masuda and K. Tanabe, *J. Appl. Phys.* **136**, 175703 (2024).