

# ニューラルネットワークを用いた化学組成による 実験値バンドギャップ予測のアンサンブル学習

## Ensemble learning for experimental band gap prediction

### from chemical compositions using neural networks

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、半導体や絶縁体を含む多様な材料の電子構造を特徴づける上で極めて重要なパラメータであり、正確な計算が求められている。しかし、標準的な密度汎関数理論(DFT)による計算では、実験値に比べてバンドギャップを約40%過小評価する傾向がある。<sup>[1]</sup>ゆえに、バンドギャップを高精度かつ高速で予測するため、機械学習を活用したバンドギャップ予測に関する研究が行われてきた。本研究では、未知の材料においては原子の構造が不明であることから、元素の組成のみからの半導体バンドギャップの予測を行うニューラルネットワークを組み合わせたアンサンブル学習モデルを提案した。

#### 計算手法

本研究では、メッセージパッシングニューラルネットワーク(MPNN)や条件付き敵対的生成ネットワーク(CGAN)といった複数のニューラルネットワークを組み合わせた実験値バンドギャップを予測するアンサンブル学習を開発した。実験データのトレーニングセットは1,986データ、テストセットは497データを使用した。アンサンブル学習の第1層モデルとして、6種類の機械学習(CGAN, MPNN, SVR, GBR, RFR, LR)とDFTによる計算データを組み合わせ、全ての組み合わせについて、予測精度を評価した。さらに、モデルの組み合わせが、バイアス、分散、Shapley値の観点から、予測結果に与える影響を分析した。

#### 結果と考察

我々が開発したモデルは、MPNNやCGANを加えることで実測値に対して平均絶対誤差0.348 eVを示し、実験値バンドギャップを予測する既存の機械学習モデルの中で最高の予測精度を達成した。このモデルは、単一SVRモデル比で12%、

従来の最良モデル<sup>[2]</sup>比で5.7%精度向上した。

Table 1 Comparison of our model with conventional models and a single SVR model.

	Single SVR	Conventional model	Our model
First stage model	-	SVR GBR RFR DFT	CGAN MPNN SVR GBR LR DFT
R <sup>2</sup>	0.858	0.883	0.894
Improvement rate [%]	-	3.0	4.3
RMSE [eV]	0.606	0.549	0.522
Improvement rate [%]	-	9.5	14
MAE [eV]	0.395	0.369	0.348
Improvement rate [%]	-	6.5	12

さらに、第1層モデルがアンサンブル予測に与える影響について、バイアス、分散、Shapley値の観点から分析した結果、CGANとMPNNを組み合わせることが予測精度向上に大きく寄与することを示した。また、CGANとMPNNを導入することでアンサンブル予測の分散は低減した一方で、バイアスが増加することがわかった。<sup>[3]</sup>

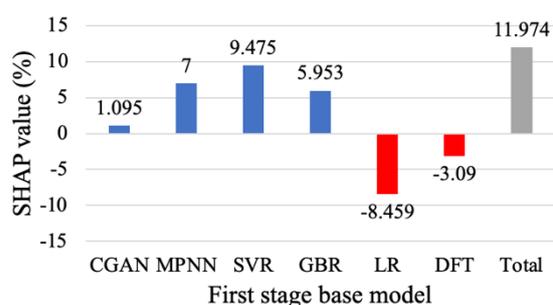


Fig. 1 Contribution analysis of first-stage models based on improvement rate over single SVR using Shapley values.

#### 参考文献

- [1] J. P. Perdew and M. Levy, *Phys. Rev. Lett.* **51**, 1884 (1983).
- [2] S. K. Kauwe, T. Welker, and T. D. Sparks, *Integr. Mater. Manuf. Innov.* **9**, 213 (2020).
- [3] T. Masuda and K. Tanabe, *Comput. Mater. Sci.* **246**, 113327 (2025).