

データ駆動型機械学習による分子構造からの物性予測と電子密度の獲得

○椿 真史¹¹産業技術総合研究所

Data-driven self-consistent learning machine for density functional theory

○Masashi Tsubaki¹¹National Institute of Advanced Industrial Science and Technology

1. はじめに

材料開発には、量子化学シミュレーションを用いた物性計算が必要不可欠であるが、膨大な計算コストがかかるという問題がある。この問題を解決するために近年、機械学習・深層学習が物性予測に用いられるようになってきた[1, 2, 3]。これらは高速かつ高精度な物性予測を実現している一方で、特に深層学習モデル内部の計算は極度に非線形かつブラックボックスであるため、材料開発の現場において重要な解釈性および信頼性が著しく低いことが大きな問題となっている。さらに機械学習技術は一般に、既存のデータから答えを導く内挿は得意だが、未知のデータに対する答えを導く外挿は不得意であり、性能が著しく悪くなる傾向が強い。例えば物性予測における外挿とは、学習データと分子量や分子構造が大きく異なる化合物の物性を予測することなどであり（内挿はその逆で、分子構造がほぼ同じ化合物の場合である）、外挿予測ができることは新規の材料開発に極めて重要である。

2. 提案モデル

本稿では、化合物の分子構造（原子配置）のみから原子化エネルギーを学習する過程において、深層モデル内部の計算が密度汎関数理論に基づくことで電子構造を獲得する、自己無撞着学習モデルを新たに提案する。このモデルではまず、与えられた分子において、原子軌道（基底関数）の線形結合、すなわち LCAO で分子（コーン・シャム）軌道を初期化する。そして、その分子軌道を入力としたニューラルネットワークを用いて、原子化エネルギーを学習する。ここで、分子軌道から得られる電子密度と原子配置から得られるポテンシャルとの対応を、もう一つのニューラルネットワークを用いて学習する。つまり、深層モデル内部の計算・学習過程に、ホーヘンベルグ・コーンの定理を

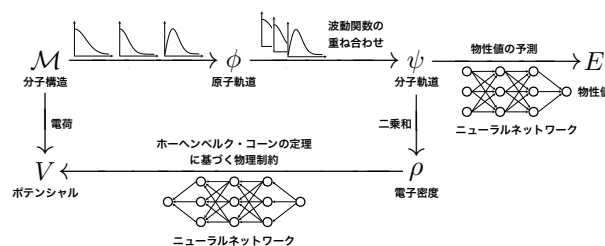


Fig. 1. 提案する深層学習モデルの全体像

物理制約として課すわけである。そして、Fig. 1 に全体像を示すように、各計算過程は一つに繋ぐことができるため、モデル全体を一貫学習できる。重要なのは、密度汎関数理論に基づくことで、モデルがデータの偏りに影響されない化合物の本質を捉え、機械学習モデルであっても外挿予測が可能になるという点である。

3. 性能評価

この提案モデルを、化合物とその物性に関する大規模データベースである QM9 を用いて学習し、物性の外挿予測性能を評価した。具体的には、理論計算は実験値を 1~2kcal/mol の誤差で予測できる一方で、提案モデルはその理論計算値を 1~3kcal/mol の誤差で予測できた；つまり、実験値を 2~5kcal/mol の誤差で予測できることになる。さらに、理論計算は 1 個の分子に数十分から数時間かかるが、提案モデルは 1 万個の分子の予測を数分で実行できる。このように、実用に耐えうる外挿予測精度を保ちながら、理論計算を 10 万倍以上高速化したこの深層学習モデルを用いることで、新規材料を効率的に探索・発見することが可能となる。

文 献

- [1] Rupp et al., Physical Review Letters 108, 058301 (2012).
- [2] Brockherde et al., Nature Communications 8, 1 (2017).
- [3] Schutt et al., The Journal of Chemical Physics 148, 241722 (2018).

*E-mail: tsubaki.masashi@aist.go.jp