

(早大先進理工¹・早大理工総研²) ○玉江 結々葉¹、中野 匡彦²、中嶋 裕也²、清野 淳司^{1,2}

Recent advances in artificial intelligence have heightened expectations for discovering novel laws and principles through data-driven automatic derivation of mathematical formulas. While symbolic regression enables the derivation of interpretable equations, it faces challenges in computational efficiency due to its simultaneous search of equation structures and coefficient optimization. This study aims to enhance the efficiency of symbolic regression by evaluating 42 optimization algorithms during the coefficient optimization phase.

【緒言】近年の人工知能（AI）技術の急速な進展により、データ駆動型の数式の自動導出法を活用した新しい法則や原理の発見に対する期待が高まっている。我々はこれまで、その手法の一つであるシンボリック回帰を化学法則の導出に適用し、その性能を検証してきた¹⁾。シンボリック回帰は、データのみから解釈可能な数式を得ることを可能にする。しかし、式の形状探索と係数最適化の両方を同時に行う必要があるため、探索空間が広大で計算負荷が高く、最適解への収束速度や局所解への捕捉などの課題が残されている。本研究では、シンボリック回帰の精度および効率を向上させることを目的として、数式の係数最適化の段階において様々な最適化アルゴリズムを適用し、ベンチマークセットを用いてその効果を検証した。

【方法】 代表的な化学法則に対して、化学的に妥当な範囲内で各変数の値をランダムに設定し、検証用データセットを準備した。既存の最適化アルゴリズム 42 種類を用い、①勾配を用いた手法、②メタヒューリスティクス手法、③ベイズ最適化手法、④その他の 4 つに分類した。各手法で得られたパラメータを用いて、真値との二乗平均平方根誤差 (RMSE) を算出した。ランダム性を考慮するために 10 回の評価を行い、精度を検証した。

【結果】本稿では、窒素分子におけるファンデルワールスの状態方程式 ($p=RT/(V-b)-a/V^2$) のパラメータ a, b, R の最適化を対象とする。図 1 に、各手法で RMSE 値が 1.0 未満となる確率を示す。この閾値は、3 つのパラメータを同時に再現するための目安となることは確認済みである。この結果、大域的探索能と局所探索能が高く、かつバランスが取れた手法(②の SA や ABC、③の GPBO+BFGS、④の PS)は、いずれも 100%を示すことが確認された。

- 1) T. Isoda, S. Takahashi, M. Nakano, Y. Nakajima, J. Seino, *J. Comput. Chem. Jpn.* **2023**, 22, 37.

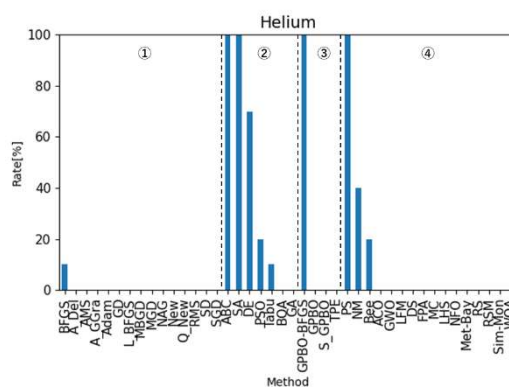


Fig 1. Performance comparison of optimization methods: proportion of runs achieving RMSE < 1.0 in fitting van der Waals equation parameters for N₂.