

分子結晶開発における機械学習の活用

(早大データ科学¹) ○谷口 卓也¹

Utilization of Machine Learning in Molecular Crystal Development (¹Center for Data Science, Waseda University) ○Takuya Taniguchi¹

In this presentation, I discuss the development of molecular crystals utilizing machine learning. Structural phase transitions, which occur upon heating and cooling, are difficult to predict, and theoretical calculations are not suitable for material screening. I utilized machine learning to screen for molecules with a high likelihood of exhibiting structural phase transitions and discovered novel crystals. Furthermore, I evaluated the prediction accuracy of elastic modulus calculations using machine learning potentials and screened molecular crystals with various Young's moduli.

Keywords : Machine learning; Molecular crystals; Elastic modulus; Structural phase transition; Materials informatics

ソフトクリスタルは分子の柔軟性、弱い分子間相互作用、結晶中の自由空間などに起因して弱い刺激で構造変化する材料群である。その多くは分子性結晶であり、多様な分子構造・結晶構造から様々な機能性を発現する。分子結晶は似た分子を使っても全く異なる結晶構造が得られることが多く、所望の性質を持つ分子結晶をねらってデザインすることは難しい。近年はデータ科学を活用したマテリアルズインフォマティクス (MI) が材料開発に広く取り入れられているが、ターゲット材料の多くは無機材料や高分子材料であり、分子結晶におけるデータ科学および機械学習の活用はそれほど進んでいない。この背景をもとに、分子結晶探索に機械学習を活用した研究について発表する。

構造相転移は、ある固体状態から別の固体相に変化する現象であり、様々な機能変化を創出できる現象である。新しい構造相転移の発見においては、分子結晶が構造相転移するかどうかを理論計算でスクリーニングすることは困難であり、実験で1つずつ

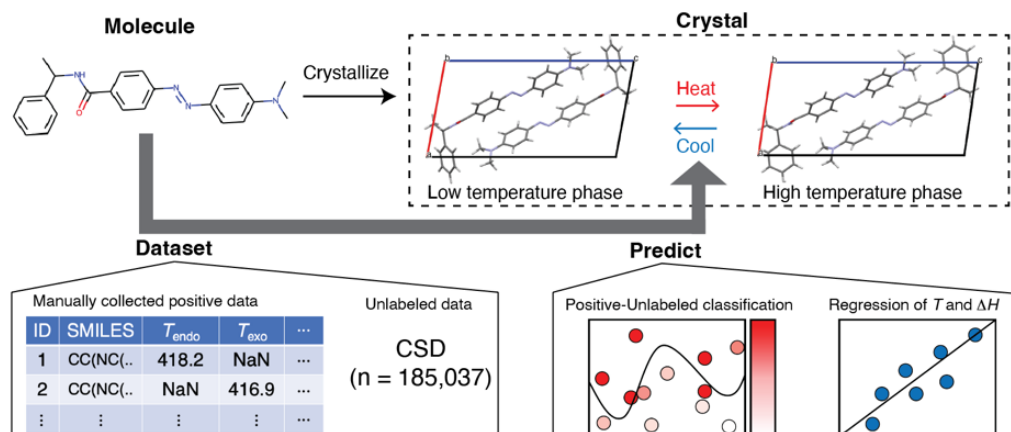


図 1. 機械学習による構造相転移の予測

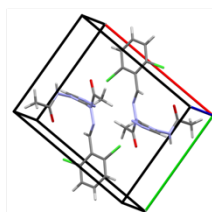
つ確認する必要がある。この工程を簡略化できれば効率的な材料創成が可能となるため、構造相転移が起こるかどうかを機械学習によって予測することとした (図 1)¹⁾。

そのために分子構造、転移温度、転移エンタルピーを学術論文から収集し、分子結晶の構造相転移の **positive** データセットを作成した。Unlabeled データは、いくつかの条件に基づいてケンブリッジ結晶構造データベース (CSD) から収集した。これらのデータで **Positive-Unlabeled** 学習を行い、どの unlabeled データが構造相転移の可能性が高いかをスクリーニングした。分子記述子として Avalon を用い、サポートベクターマシンを予測関数とした場合が最も良い分類器であった。この分類器でスクリーニングした結果、構造相転移の可能性が高いと予測された候補分子のうち約 8% で実際に相転移が確認された。これは、CSD に含まれる相転移の報告率約 0.3% を大きく上回る数値であり、従来の方法ではできなかった大規模スクリーニングである。

次に、分子結晶におけるグラフニューラルネットワーク (GNN) の有効性を検証した²⁾。GNN は近年 MI で注目されている手法であるが、分子結晶においては分子グラフと結晶グラフの両方のデータ表現が可能である。タスクをバンドギャップ予測に設定し、両者のグラフ表現を比較した結果、予測誤差は分子グラフよりも結晶グラフの方が小さいことがわかった。この結果自体は当然とも言えるが、結晶グラフの予測誤差は分子グラフの予測誤差の約 0.4 倍になると定量評価できた。また、CSD から抽出した大規模データセットのバンドギャップをスクリーニングでき、大小それぞれのバンドギャップを持つ可能性のある結晶を特定できた。

さらに、分子結晶の弾性率予測におけるニューラルネットワークポテンシャル (NNP) の有効性を検証した³⁾。弾性特性は、低分子医薬品の錠剤化や材料応用にとって重要だが、理論計算で弾性定数テンソルを計算することは計算コストが高い。理論計算と比較して計算コストが低いことで注目される NNP を使用して分子結晶の弾性率

Molecular Crystals



Elastic Modulus

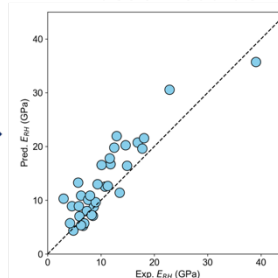


図 2. ニューラルネットワークポテンシャルによる分子結晶の弾性率予測

を予測した結果、計算された弾性率は実験値と十分に一致した (図 2)。また、NNP による計算でナノインデンテーションの測定結果も概ね再現でき、ヤング率の大きさと結晶構造との関係も見出すことができた。さらに、計算コストの低さを活かして分子結晶の弾性率をスクリーニングし、5000 個の結晶構造の弾性率を算出し、大きなおよび小さな弾性率を持つ結晶を見出した。

- 1) D. Takagi, K. Ishizaki, T. Asahi, T. Taniguchi, *Digital Discovery*, **2023**, 2, 1126-1133.
- 2) T. Taniguchi, M. Hosokawa, T. Asahi, *ACS Omega*, **2023**, 8, 39481-39489.
- 3) T. Taniguchi, *CrystEngComm*, **2024**, accepted.