

実験と理論・計算・データ科学を融合した材料開発の革新
2019年度採択研究代表者

| |
|-----------------|
| 2019年度 実績報告書 |
|-----------------|

内藤 昌信

物質・材料研究機構統合型材料開発・情報基盤部門
グループリーダー

データ駆動型分子設計を基点とする超複合材料の開発

§ 1. 研究成果の概要

(1) 機械学習による分子材料および有機反応探索

マテリアルインフォマティクスを用いて、新規な有用(高)分子群を提案するため、世界最大の高分子データベース PoLyInfo と独自の機械学習アルゴリズム(SMILES-X)の開発を行った。これまでの化学物質を探索するための機械学習アルゴリズムでは、教師データとなる化合物をフィンガープリント化する必要があった。一方、本プロジェクトで開発したアルゴリズム“SMILES-X”では、分子構造情報(smiles)以外の記述子を用いることなく材料特性を予測できることを特徴とする。これにより、高い予測精度を維持したまま、大幅に計算コストを削減することに成功した。例えば、1128種の有機分子における水溶解度データベースを用いて(Delaney J S 2004J. Chem. Inf. Comp. Sci.441000-5)、分子構造のみから水に対する溶解度を予測すると、その平均平方二乗誤差(RMSE)は 0.57 ± 0.07 (mols/L)と、過去の研究を通して非常に高精度な予測ができた。また、本予測を行った際に、SMILES中のどのトークンが重要であったかを描画することで、現象の解釈・分析を行うことが可能となった。

(2) スマートラボとIoT化

本プロジェクトでは、高分子・複合材料の合成・化学分析・力学物性評価を自動化し、得られた良質のメタデータをもとにデータベースを迅速に構築することで、機械学習の予測精度を向上させることを目指している。そのための手法として、実験で得られるメタデータを物質・材料研究機構のデータプラットフォームに蓄積し、自動解析するプログラムを作成した。このシステムを様々な計測装置群に展開していくことで、本プロジェクトで得られる生データを効率的にデータベース化することが可能となり、今後のデータ駆動による材料開発が大幅に加速化することが期待できる。

一例として、ヒトと人工知能の協働による機械学習“アクティブラーニング”による高強度接着剤の開発を行った。機械学習による条件判断の正確さを活用することで、少ない実験数にも関わらず、世界最高強度レベルの構造用接着剤を開発することに成功した。

(3) 力学特性評価(ハイスループット化)の作製とデータベース化の基盤構築

本プロジェクトでは、得られた高分子・複合材料について、化学分析のみならず、材料としての特性評価をハイスループット化することを目指している。本年は、力学特性の自動測定装置の作製と、得られたデータをデータベースに取り込むシステム構築を行った。一例として、高分子の接着性評価を目的としたハイスループットせん断強度測定装置を開発した。

【代表的な原著論文】

1. Guillaume Lambard, Ekaterina Gracheva, “SMILES-X: autonomous molecular characterization for small datasets without descriptors”, Mach. Learn.: Sci. Technol., vol. 1, No. 2, 025004, 2020

§ 2. 研究実施体制

(1) 超複合材料開発グループ

- ① 研究代表者:内藤 昌信 (物質・材料研究機構統合型材料開発・情報基盤部門 グループリーダー)
- ② 研究項目
 - ・データ駆動型分子設計を基点とする超複合材料の開発

(2) 分子デザイングループ

- ① 主たる共同研究者:袖山 慶太郎 (物質・材料研究機構統合型材料開発・情報基盤部門グループリーダー)
- ② 研究項目
 - ・機械学習による分子設計・反応予測モデルの開発

(2) 材料特性評価グループ

- ① 主たる共同研究者:佐藤 千明 (東京工業大学科学技術創成研究院 教授)
- ② 研究項目
 - ・材料特性評価の自動データ収集システムの構築