

ポリマー探索に AI を活用してみよう

内藤昌信

物質・材料研究機構 高分子・バイオ材料研究センター
 [305-0047] 茨城県つくば市千現1-2-1
 副センター長, 博士 (工学).
 専門は高分子材料科学.
 NAITO.Masanobu@nims.go.jp
https://www.nims.go.jp/group/Data-driven_Polymer_Design/index.html

新たな機能性ポリマーを探索する手段として、人工知能 (AI) を活用するポリマー・インフォマティクスが注目されている。本稿では、ポリマー・インフォマティクスに求められる、情報科学、データベース、AI ツール、自動合成・評価・プロセス技術、オートメーション化などを俯瞰するとともに、最近の著者らの取り組みを紹介する。

1. ポリマー・インフォマティクス

1.1 AI 活用に必要なアイテム

AI を活用したポリマー・インフォマティクスを進める上で、必要となるアイテムとその相関を図1にまとめた。言うまでもなく、AI 活用の際に最も重要なのが高分子の合成、構造、物性などに関するデータセットである (図1①)¹⁾。これらデータベースにアクセスできるならば、まずは欠損データの補完などデータを AI が読めるようなフォーマットへと前処理する。機械可読化となったデータセットをもとに回帰モデルや深層学習モデルなど、与えられた課題を解くために必要なアルゴリズムを選択することで、最適な実験条件の提案、新規分子の生成や予測などを AI が行えるようになる (図1②③)。予測精度の向上のためには、データベースの拡充が重要であるが、AI と研究者が協働しながら実験データを拡充していくアクティブラーニング (能動学習) は、効率的にデータを取得するために有力な手法である²⁾。アクティブラーニングは、AI が判断しにくい境界領域を専門家が実験データをピンポイントで加えていくことで予測精度を上げる方法である。また、実験のハイスループット化や省人化も

注目されている (図1⑤)。最終的には、ポリマー・インフォマティクスを構成する各要素技術をクロズド・ループとして循環させながら探索空間を広げていくことで未知のポリマー探索に繋げていくことが、ポリマー・インフォマティクスには期待されている。

1.2 高分子データベース

AI の活用には、何はなくともデータベースが欠かせない。ところが、無機材料などと異なり、高分子材料のデータベースは、質・量ともに少なく、また、種類や物性値が包括的にカバーされていないのが現状である (SI)。例えば、世界最大の高分子データベースである PoLyInfo は、論文に記載されているデータを抽出・収集していることもあり、熱特性に関連するデータの約30%が10種類のポリマーのものであり、さらにその中の40%がガラス転移温度のデータセットであると報告されている³⁾。このようなポリマー・インフォマティクスが抱える初期データセットの壁を克服するため、計算科学の自動化によるデータ生成や⁴⁾、AI とスマートラボの協働による能動学習⁵⁾などが進められている。

1.3 スマートラボと機械学習ツール

ポリマー探索に AI を活用するためには、合成・構造・物性・材料特性を含む幅広いデータベースが求められる。しかし、ターゲットとする材料の性能や特性は目的により刻々と変化する。そのため、スマートラボを用いたデータベース構築には、臨機応変に対応できるロバスト性が求められる。著者らが開発を進めるポリマー・スマートラボでは、これらの課題解決のため、技術仕様を予め固めずに様々な AI ツールやデータサービスを組み合わせるアジャイル的な開発を採用している。その根幹となる AI ツールやデータベースには、物質・材料研究機構 (NIMS) が開発を進める材料プラットフォーム「DICE」を最大限に活用した。NIMS は、文部科学省「マテリアル DX プラットフォーム構想」におけるデータ中核拠点として、産学の高品質なマテリアルデータの戦略的な収集・蓄積・流

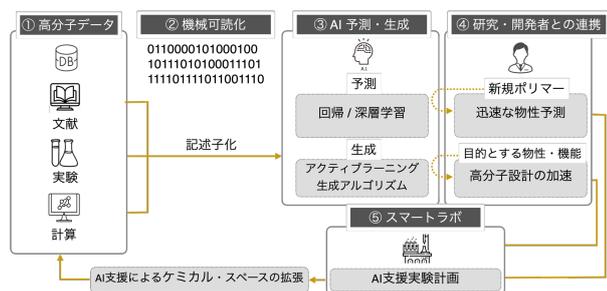


図1 ポリマー・インフォマティクスの要素技術

通・利活用を進めている。その基盤となる材料プラットフォームである「DICE」は、材料データの入り口(集まる)から出口(使う)までを一貫して扱う総合的な機械学習ツール群である。データ収集の入口には、実験・計測装置や文献からの取り込み、機関レボジトリからの取り込みなど、複数の情報源からのハイスループット・ハイクオリティー収集・登録システムを実装するとともに、データをRDF形式で階層的、構造的に理解し、かつLinked Dataなど情報工学的手法によってデータ間を繋ぐことで、格納データの高付加価値化に繋げている。またデータの利活用になる出口側では、データ可視化システムや各種解析ソフトウェアを整備・提供し、データ駆動型材料開発における様々な課題解決の糸口を与えられるような、システム科学的なサービス提供を行っている。著者等のポリマー・スマートラボの開発においても、データ解析AI、データサーバなど基本的な機械学習モジュールをDICEが提供する各種サービスを利用することで、システム開発の迅速化・低コスト化を図っている(図2)。中でも、DICEツールの一つであるRDE

(Research Data Express)は、物質・材料に係るデータを保管し、国内の物質・材料研究に供するためにデータをクラウド上で管理するシステムである。主なサービスは、研究データ蓄積・管理サービス、研究データ公開、共用(検索・閲覧)サービス、認知情報、装置情報、試料情報、語彙(装置・試料に関する用語)などのマスタ管理である。例えば、RDEに生データを登録すると、構造化されたデータをクラウド上にセキュアな状態で保存することができる。また、機械学習を行う際のトラブルの主要原因となるデータの前処理にかかる時間も最小限で済ますことができる。これによりユーザーや研究グループ内での再利用、また、他の研究グループとのデータの共用が容易となり、マテリアル研究開発のDX化がシームレスに行えるようになった。このRDEは文部科学省が推進するデータ創出・活用型マテリアル研究開発プロジェクト(DxMT)やマテリアル先端リサーチインフラ(ARIM)においてもデータ活用のためのツールとして採用されており、プロジェクトを通じて創出されたデータを、機械学習に適した形で保管・運用する試みが始まっている。

ポリマー・スマートラボの基本構成にはDICEの機能を最大限に活用しつつ、一方で、高分子材料の自動合成装置や、在庫管理システムなど、高分子材料研究に特有のモジュールや機械学習ツールについては、自前で開発を進めることでトータルの研究開発のスピードとコストの最適化を図っているのが著者等が開発を進めるポリマー・スマートラボの特徴である。例えば、有機・高分子材料のハイスループット化は3Dプリンタの登場により自動合成装置のツールを研究室で自作することも可能になってきた。また、ロボット実験装

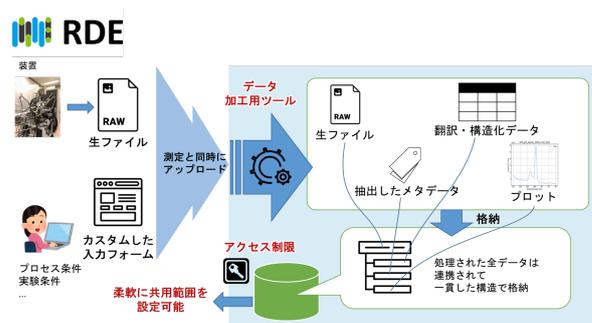


図2 スマートラボのアジャイルシステム開発。RDEは実験データをオンラインで迅速に登録するためにNIMSが開発したマテリアルズ・インフォマティクスのための機械学習ツール

置と材料探索用人工知能(AI)を人が介入することなく連携させ、自律自動材料探索を可能とするための汎用ソフトウェア、NIMS-OS(NIMS Orchestration System)が開発され、オープンソースソフトウェアとして公開されている⁵⁾。このように合成に関するハイスループット化の進捗に対し、熱物性や力学物性の評価はサンプル調製などに多くの時間・労力・職人技などが求められることから、ハイスループット化が難しいという課題が残っている。その解決策として、著者らはデータ駆動型研究に特化した材料評価ハイスループット装置を開発し、合成から材料評価までの一連の工程を自動化・高速化する試みを行っている。自作した力学特性評価や粘弾性特性評価などのハイスループット化することで、接着材料^{2,6)}や力学傾斜材料⁷⁾などの新たな機能性高分子材料の開発に威力を発揮している。

1.4 ポリマー・シークエンサー

AIによる材料探索から実用化までをシームレスに繋ぐためには、マテリアルズ・インフォマティクスが得意とする新規材料の探索と、製品化に必要なプロセス・インフォマティクスを連携させることが不可欠である。ところが、マテリアル・インフォマティクスに用いられる分子記述子には、化学構造や組成情報は含まれているものの、プロセスの影響を大きく受ける高次な構造情報は含まれていない。一方で、プロセス・インフォマティクスでは、高分子材料の成形加工の過程で生じる物性や機能が重要になってくることから、高次構造に由来する情報を含む記述子が必要となる。材料特性データを取得するためには、固体状態の材料を溶剤に溶解させたり、表面研磨といった前処理が必要となることがしばしばある。また、多くの実材料にはさまざまなフィラーや添加剤が含まれており、高分子材料の評価分析を困難にする要因となっている。そこで著者らは、プロセスを反映した高次構造の情報をできるだけ保持したまま、MIに必要な化学情報を記

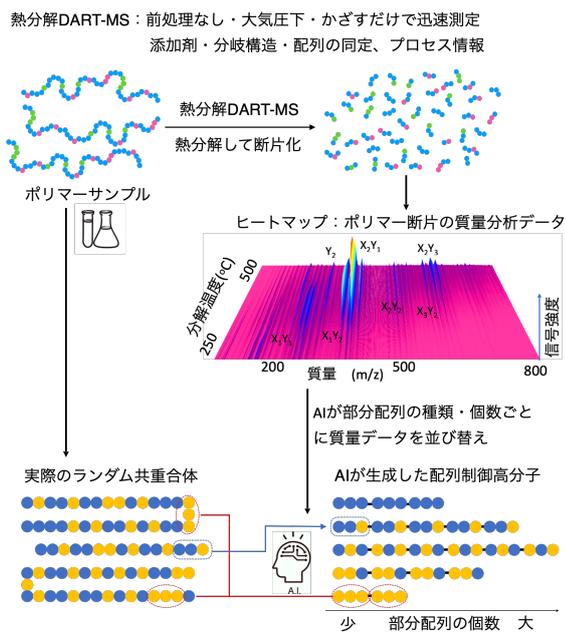


図3 直接イオン化質量分析 (DART-MS) を用いた高分子鎖の仮想的再配列による部分配列の定量化

述するための AI 解析手法として、熱分解 (Pyrolysis) により得られる大量の質量スペクトルを記述子として用いる手法を見出した。本来、質量分析データを定量的に解析するためには標品 (系の構成成分の純品) が不可欠であったが、大量の質量分析データに「教師なし学習」の一種である非負値行列因子分解 (NMF) を適用することで、リファレンスフリー定量質量分析 (Reference-free quantitative MS; RQMS) を可能とした。具体的には、直接イオン化法質量分析 (DART-MS) に取り付けたパイロライザーで室温から 600°C までサンプルを加熱すると、熱分解しやすいところからサンプルが断片化していく。得られた大量の質量スペクトルデータを AI 解析することで、質量スペクトルをリファレンスフリーで定量化することに成功した (図3)⁸⁾。RQMS を用いた例として、3種類の高分子材料の混合物の組成比の同定結果を示す (図4)。標品の測定を行わずとも、AI が組成比の頂点 (各ポリマー単独成分) を導き出している。また、DART-MS の実測値と RQMS による予測値の差は極めて小さいことから、主成分となる樹脂の中に含まれる微量成分の同定にも威力を発揮することがわかる。実際、この AI によって再構成された質量分析データを記述子として用いることで、樹脂成分に含まれる数 100ppm オーダーの不純物成分の検出やフィラーや添加剤を含む高分子材料の構造解析が可能となってきた。従来の分析技術では実現できない物性や構造情報を AI を活用することで明らかにすることができた好事例といえよう。

2. おわりに

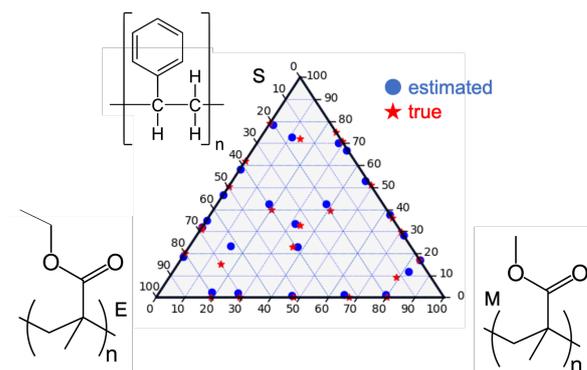


図4 直接イオン化質量分析 (DART-MS) を用いた高分子鎖の仮想的再配列による部分配列の定量化 青丸: RQMS による予測値、赤星: DART-MS による実測値

本稿では、AI でポリマーを探索するためのはじめの一步に必要なツールや技術について紹介した。一般的に新材料の開発期間には 10~20 年かかると言われる。その中で、材料開発の競争に世界で勝っていくためには、AI の活用が不可避であることは言うまでもない。その中で、我が国は PoLyInfo に代表される高分子データベースや、マテリアル研究 DX プラットフォームを通じた実験データの収集・保管、さらには、得られた良質な材料データベースをオープン・クローズド戦略に基づいて運用するためのノウハウが培われている。これは、分かち合いの価値観や三方良しといった日本の伝統的な価値観に基づいたものとも言える。特に、これからの AI 活用によるポリマー探索においては、データを戦略資源と捉え、積極的に活用していくことが必要と思われる。

謝 辞

本研究は JST-CREST (JPMJCR19J3) の支援を受けて行われた。本稿で紹介した研究成果は、物質・材料研究機構高分子・バイオ材料研究センター中村泰之主任研究員、日比裕理研究員、藤田健弘博士をはじめとする NIMS データ駆動高分子設計グループメンバーのご尽力によるものである。誌面を借りて謝意を顕します。

文 献

- 1) L.Chen, *et al.*, *Mater. Sci. Eng. R: Reports* **144**, 1 (2021)
- 2) S.Pruksawan, *et al.*, *Sci. Technol. Adv. Mater.* **20**, 1010 (2019)
- 3) ウステファン *et al.*, *統計数理* **69**, 65 (2021)
- 4) Y. Hayashi, *et al.*, *npj Computational Materials*, **8**, 222 (2022)
- 5) R. Tamura, *et al.*, *Sci. Technol. Adv. Mater. Methods*, **3**, 2232297 (2023)
- 6) C. Kang *et al.*, *J. Adhes.* **99**, 2080 (2023)
- 7) W.-H. Hu, *et al.*, *Sci. Technol. Adv. Mater.* **23**, 66 (2022)
- 8) Y.Hibi, *et al.*, *Chem. Sci.* **14**, 5619 (2023)

-----本文ここまで-----

Supporting Information

表 1 一般公開されている高分子に関連するデータベース

<p>PoLyInfo https://polymer.nims.go.jp</p>	<p>物質・材料研究機構が運営している高分子物性データベース。学術論文 (20, 959 報) から抽出した物性値等を収録。物性値のデータポイント総数 519, 175 2024 年 4 月 14 日現在</p>
<p>Polymer Genome https://www.polymergenome.org</p>	<p>文献および DFT 計算で求めた物性値を提供。電子物性、力学物性、熱物性、溶解性など 22 種類の物性値が収録されている。</p>
<p>MaterialsMine http://www.materialsmine.org</p>	<p>ポリマーコンポジットの電子顕微鏡画像や物性データベース (Nanomine) とメカニカルメタマテリアルのシミュレーションツール等の提供 (Metamime) および各種実験データの</p>

	レポジトリ
<p>Citrination https://citrination.com</p>	<p>ポリマーをはじめ、マテリアルデータを公開するためのプラットフォーム</p>
<p>CAMPUS https://www.campusplastics.com</p>	<p>市販されているポリマーの材料特性データベース</p>
<p>MATWEB Material Property Data http://www.matweb.com</p>	<p>市販されているポリマーの材料特性データベース</p>
<p>Material Properties Database https://www.makeitfrom.com</p>	<p>ポリマーを含むエンジニアリング材料の特性比較データベース</p>
<p>Polymer Property Predictors and Database https://www.nist.gov/projects/polymer-property-predictor-and-database</p>	<p>NIST が公開している高分子データベース。文献より抽出した Flory-Huggins χ パラメータやガラス転移温度が収録。</p>