

マテリアルズインフォマティクスの活用による 機能性・芳香族高分子の探索と実証

(東工大 物質理工) [○] 畠山 敏, 石川 弘樹, 難波 江裕太, 早川 晃鏡

【要旨】

マテリアルズインフォマティクス (MI) の進展や展望について、イオン伝導性を示す高分子固体電解質の予測モデルやポリイミド前駆体の自動合成装置の動作状況を例示しながら紹介する。また、GPT-4 などの大規模言語モデルの化学研究における現状や展望について議論する。

【緒言】

マテリアルズインフォマティクス (MI) は材料科学の事象をデータサイエンスの視点で解析する分野である。近年は深層学習、高度シミュレーション、ロボティクスとの融合も進み、研究ツールとしての有用度が高まってきた¹。2023 年は革新的な大規模言語モデル GPT-4 が登場し、AI との向き合い方が一変した。このような状況下では、MI の現在地を改めて俯瞰する作業が必要になる。そこで本発表では、機能性高分子や芳香族ポリマー研究における最近の MI の適用事例について紹介しながら、当該分野の課題や展望を議論する。

構造—物性相関予測モデルを用いた芳香族系イオン伝導体の探索

分子構造と実験結果の関係を包括的に記録したデータベースの整備は、科学法則の抽出や物性予測のための機械学習モデルの開発に役立つ。このアプローチを用い、リチウムイオン伝導性を示す高分子組成物の分子構造から伝導度を推定するモデルを構築した²。初めに公開論文から約 1 万件のリチウム伝導性ポリマーや有機電解質のデータを手動で収集し、データベースを作成した (図 1 左参照)。続いてグラフニューラルネットワークやガウス過程のような機械学習アルゴリズムを採用し、電解質の情報 (分子構造、組成比、計測温度) からイオン伝導度を予測するモデルを開発した。この一連のデータベースと解析スクリプトはウェブ上で公開している²。

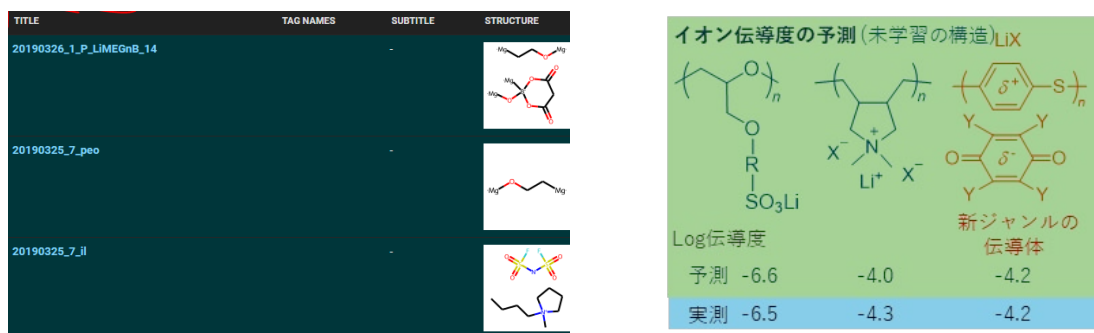


Fig. 1 左: 電解質データベースの様子. 右: 電解質の伝導度予測の例

本研究で構築した予測アルゴリズムは、未知の電解質構造に対しても優れた予測性能を示した。例えばスルホ基を有するポリエーテル系リチウムイオン電解質の室温での実測イオン伝導度が Log スケールで-6.5 であったのに対し、予測値はほぼ同等の-6.6 だった。アンモニウム系高分子電解質においても、予測値(-4.3)と実測値(-4.0)が概ね一致した。これらの精度は熟練の電解質研究者の予測に匹敵するレベルである。

構築した予測モデルは新規な分子設計に基づく電解質の探索にも有効であった。従来あまり注目されてこなかった芳香族系ポリマーの組成物についても、実測値と予測値が一致するケースを観測した。当該ポリマーのガラス転移点は 100℃以上と高かったにも関わらず、電解質に用いた際の伝導度は液体レベル(>0.1 mS/cm)であった。本事例のように、メカニズムが一見不明な新規の電解質系に対しても回帰モデルが優れた予測性能を示した点は意義深い。追加の実験研究により、芳香族ポリマーからなる電解質系でイオン輸送を担うのは、電解質内でアモルファス構造化したリチウム塩³であることも明らかになり、新たな研究展開に繋がっている。

大規模言語モデルやロボティクスの活用

ここ数年の人工知能分野においては、従来の専門特化型の AI から、汎用性の高い基盤モデルの活用というパラダイムシフトが起きつつある。これまでの AI は、「知能」というよりはむしろ、特定のタスクを解くために専用のアルゴリズムというイメージが強かった。例えば深層学習を活用した顔認識アルゴリズムや将棋・囲碁のプログラムは人間を遥かに凌駕した性能を示すが、そこから人間的な知性を読み取る者は少ない。

一方、膨大な量のテキストを学習した GPT-4 のような大規模言語モデルは、幻覚などの問題はまだ残っているものの、「人工知能」を彷彿とさせる汎用性や推論性能を示し始めている。2024 年 1 月には GPT-4 が大学共通テストの大半の科目で受験者平均を上回り⁴、同様の機構で動くモデルが数学オリンピックで金メダル級の実力⁵を示したことが話題となった。今後の化学分野においても、大規模言語モデルの活用が加速する可能性が高い⁶⁻⁸。

大規模言語モデルとは対照的に、これまで MI で活用されてきた特化型 AI はその性質上、化学分野における極めて限られたトピック—例えば特定の化合物系での構造-物性相関—



Fig. 2 AI がカバーできる知識分野の概念図

しか学習できなかった(Fig.2)。AI が限られた知識領域しか学習できないという制約は様々な副作用もたらす。例えば先述のイオン伝導体の回帰モデルは「メタンとリチウム塩を混ぜると、常温常圧で 10^{-7} S/cm のイオン伝導度を発現する」という予測を出力してしまう。しかしメタンはガスなので、固体電解質の媒体としては当然使えない。このような初歩的なミス AI が犯すのは、予測モデルにとっては、事前に学習した「1 万件のリチウムイオン伝導体」が知識の全てであり、「世の中にはイオン伝導性を持たない化合物が存在する」という常識的な事実を学んでいないためである(Fig.2)。同様に、既存の AI システムによって提案される分子構造が合成難易度や安定性、実用性等の面で多くの課

題を含みやすい傾向は、AIが「化学の常識」を持たない点に由来している¹。

これに対し、大規模言語モデルは従来の専用・特化型 AI の課題を解決する潜在力を持っている。これらのモデルは世界の多様な事象を事前学習しており、常識的かつ化学的な推論を行う能力を有しているからである。その活用研究はまだ始まったばかりだが、言語モデルを使用することで、従来の手法よりも高い解釈性と信頼性のある化学予測や提案が可能になりつつある。例えば反応条件の探索タスクにおいて、既存の化学理論を考慮しないモデルよりも優れた探索性能を示す事例もあり、分子物性の予測タスクにおいても、どのような分子構造や理論に着目すべきかについて GPT-4 が有意義な提案を行うことができています^{6,8}。ただし当該モデルは最先端の科学知識はまだ有していないため、論文誌などを追加で学習させる基礎研究⁹が必要な段階である。

言語モデルはプログラムコード生成などのタスクにも適しており、GPT-4 を利用して実験操作を行うロボットアームの制御プログラムを自動生成できる。この技術の活用を通して、AI ロボットによる自律的な化学実験システムの構築が進む可能性がある^{6,10,11}。例えば自然言語によるロボットハンドの操作を指示できるシステムが提案された¹²。将来的には手作業で行われてきた実験操作の自動化が進み、人手不足や安全性の懸念が解消されるかもしれない。

筆者らもポリイミド前駆体の合成実験を開始した。例えば 3D プリンタを応用した自家製のピペティング装置の AI 駆動の制御に取り組んでいる。このシステムには深層学習モデルを用いた画像認識機能や GPT が組み込まれているので、日本語で指示を行うだけで、ピペット操作を自動で行える。

構築したシステムを用いて、ポリイミド前駆体の合成検討を行った。この過程では、二種類のモノマーを混合し、一定時間攪拌した後、シャーレ上で反応液を加熱乾燥させる処理が自動で行われる。このシステムは「何時何分何秒に試薬を〇〇mL 移動した」といった詳細な情報を記

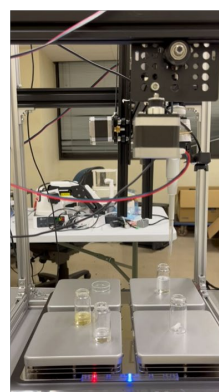
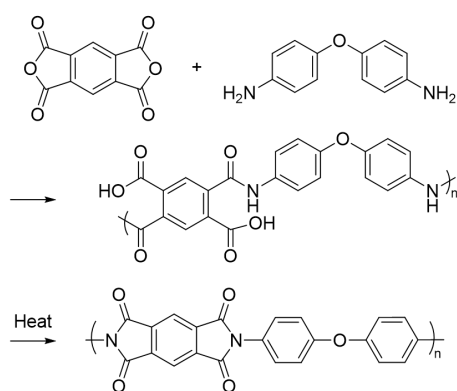


Fig. 3 ポリイミドの合成スキームと自動合成装置の様子

録でき、実験の再現性を保証し、属人性を排除する上で有益なことが判明した。ただし、現時点のプロトタイプは大気条件での実験操作しか行えないので、反応液の吸湿による重合阻害の問題が生じている。今後は不活性条件での反応や湿度モニタリング等を基軸にした、より高度な合成実験を計画している。

【結論】

マテリアルズインフォマティクス (MI) の進展と応用例、特に機能性高分子や芳香族ポリマー研究への応用に焦点を当てて議論した。MI はデータサイエンス技術と材料科学の知識を統合し、新材料の探索や既存材料の特性解析を効率化する。本研究で示した芳香族系イオン伝導体の探索や予測モデルの構築は、MI の潜在能力を具体的に示す事例といえる。また、GPT-4 などの大規模言語モデルの進化により、従来困難とされてい

た分野へのアプローチが可能になってきた。この技術は科学的知見の集約や新研究仮説の生成に貢献するので、今後は AI ロボットによる自動化化学実験システムの開発を通して、実験の再現性向上や安全性の確保がしやすくなるメリットがある。

今後の展望として、大規模言語モデルやロボティクス技術の発展により、高度で効率的な材料開発が期待される。これらの技術を利用したデータ駆動型アプローチは、材料科学だけでなく他の科学分野においても新たな研究方法論を提供する可能性がある。

【謝辞】

本研究は JSPS 科研費 21H02017、JST 創発的研究事業 JPMJFR213V 助成を受けたものです。また、自動実験装置は東北大学 高石 慎也先生に作製していただきました。深く感謝申し上げます。

【参考文献】

- [1] K. Hatakeyama-Sato, *Polym. J.*, 2022, **55**, 117-131.
- [2] K. Hatakeyama-Sato, T. Tezuka, M. Umeki and K. Oyaizu, *J. Am. Chem. Soc.*, 2020, **142**, 3301-3305.
- [3] K. Hatakeyama-Sato, M. Umeki, H. Adachi, N. Kuwata, G. Hasegawa and K. Oyaizu, *Npj Comput. Mater.*, 2022, **8**, 170.
- [4] <https://note.com/lifeprompt/n/n87f4d5510100>
- [5] T. H. Trinh, Y. Wu, Q. V. Le, H. He and T. Luong, *Nature*, 2024, **625**, 476-482.
- [6] K. Hatakeyama-Sato, N. Yamane, Y. Igarashi, Y. Nabae and T. Hayakawa, *Sci. Technol. Adv. Mater.: Methods*, 2023, **3**, 2260300.
- [7] K. Hatakeyama-Sato, S. Watanabe, N. Yamane, Y. Igarashi and K. Oyaizu, *Digital Discovery*, 2023, **2**, 1548-1557.
- [8] D. A. Boiko, R. MacKnight, B. Kline and G. Gomes, *Nature*, 2023, **624**, 570-578.
- [9] K. Hatakeyama-Sato, Y. Igarashi, S. Katakami, Y. Nabae and T. Hayakawa, 2023, **arXiv:2312.03360**.
- [10] N. J. Szymanski, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, M. J. McDermott, M. Gallant, E. D. Cubuk, A. Merchant, H. Kim, A. Jain, C. J. Bartel, K. Persson, Y. Zeng and G. Ceder, *Nature*, 2023, **624**, 86-91.
- [11] G. Tom, S. P. Schmid, S. G. Baird, Y. Cao, K. Darvish, H. Hao, S. Lo, S. Pablo-García, E. M. Rajaonson, M. Skreta, N. Yoshikawa, S. Corapi, G. D. Akkoc, F. Strieth-Kalthoff, M. Seifrid and A. Aspuru-Guzik, 2024, **chemrxiv-2024-rj946**.
- [12] <https://deepmind.google/discover/blog/rt-2-new-model-translates-vision-and-language-into-action/>