

シンポジウム | 中長期テーマシンポジウム：マテリアルズ・インフォマティクスが拓く有機機能材料

2024年3月19日(火) 13:00 ~ 15:40 皿 G02(階段教室 小)

[G02-2pm] マテリアルズ・インフォマティクスが拓く有機機能材料

座長、シンポジウム関係者：安田 琢磨、佐伯 昭紀、田村 宏之

多量の計算・実験データに基づくマテリアルズ・インフォマティクス (MI) による有機・無機機能材料の開発研究に期待が集まっている。特に多彩な化学構造と集合体構造を示す有機材料 (高分子、分子) では、発光、触媒、光電変換、熱伝導、機械的強度といった機能を最大限発揮させるために多くの検討が必要なため、MIによるアプローチが有効である。本企画では、各分野の計算、実験研究において先端を走っている研究者を講演者として招き、今後の化学分野におけるMIの発展に向けた課題を議論する。

◆ 日本語

13:00 ~ 13:30

[G02-2pm-01]

大規模言語モデルやロボット装置を活用した高分子材料の合成検討と展望

○畠山 歓¹、石川 弘樹¹、高石 慎也²、難波江 裕太¹、早川 晃鏡¹ (1. 東京工業大学、2. 東北大学)

◆ 日本語

13:30 ~ 14:00

[G02-2pm-02]

小規模データのマテリアルズインフォマティクスで挑む有機機能材料の探索と性能向上

○緒明 佑哉¹ (1. 慶大)

◆ 日本語

14:00 ~ 14:30

[G02-2pm-03]

自動分子シミュレーションによる高分子物性データプラットフォームの産学共創

○林 慶浩¹ (1. 統数研)

14:30 ~ 14:40

休憩

◆ 日本語

14:40 ~ 15:10

[G02-2pm-04]

量子化学計算と機械学習を活用した有機EL材料の開発

○相澤 直矢¹ (1. 大阪大学)

◆ 日本語

15:10 ~ 15:40

[G02-2pm-05]

機能性ソフトマテリアルの構造・物性解明: 計算科学とデータ科学によるアプローチ

○渡辺 豪^{1,2} (1. 北里大、2. 神奈川県産総研)

大規模言語モデルやロボット装置を活用した高分子材料の合成検討と展望

(東京工業大学 物質理工学院¹・東北大学 大学院理学研究科²)

○畠山 歓¹、石川 弘樹¹、高石 慎也²、難波江 裕太¹、早川 晃鏡¹

Synthesis of polymeric materials assisted by large-scale language models and robotic devices
(¹Tokyo Institute of Technology, ²Tohoku University)

Kan Hatakeyama¹, Hiroki Ishikawa¹, Shinya Takaishi², Yuta Nabae¹, Teruaki Hayakawa¹

Automating chemical research using deep learning models, such as GPT-4, is gaining attention. One of its applications is collaboration with robotics, where precise control and acceleration of material processes are expected. This presentation will introduce its characteristics via polyimide synthesis. Additionally, we will discuss the status and challenges of large language models and laboratory automation.

Keywords : *Materials informatics, Deep learning, Robotics*

背景: 大規模言語モデルやラボラトリーオートメーションに関する研究動向

[大規模言語モデル]

ChatGPT(22年11月)を皮切りに、大規模言語モデル(Large Language Model, LLM)の世界的なブームが起きている¹。GPTが人間を代替する思考能力を持っているとまだ言えないが、外国語の読解・作文やプログラミングのような規則・経験則が通用しやすいタスクでは定番のツールになりつつある。

今後の LLM 構築研究の大きな方向性として、1) より高性能なモデル、具体的には汎用人工知能(AGI)も見据えた拡大路線、または2) GPT-3 (2020)前後のサイズを持つモデルの作り込みや用途特化が挙げられる¹。OpenAI や Google などの莫大な資金力を持つ研究グループ以外は総じて後者に注力しており、日本でも GPT-3 相当のモデルを作る国家プロジェクトの公募が始まり、24年以降に推進される見込みである。

化学反応や実験条件の理解・提案などは科学法則や経験則に基づいて概ね可能なので、適切なデータセットを学習させれば、GPT-3 相当のモデルでも熟練者に近い意思決定を行える可能性はある。しかし化学分野は学習に適したプレプリントやオープンアクセス論文が少ない上、少なくとも本国では LLM を研究推進する主要コミュニティの主な興味に化学が含まれていない雰囲気²がある。まずは化学と LLM 融合分野の現在地を明確化する作業^{3,4}が必要かもしれない。

[ロボティクスとラボオートメーション]

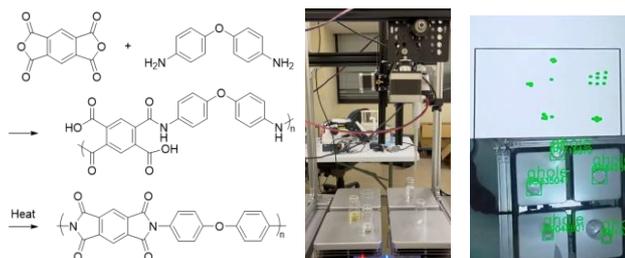
LLM と相性が良好な分野として、ロボティクスやラボラトリーオートメーションが挙げられる。ロボット制御のためには、人間であれば無意識に行える知覚・判断をプログラムコードとして逐一記述する必要があり、実装コストが高かった。一方で最近では画像認識能力を持つ LLM 活用し、自然言語で汎用ロボットアームを制御する手段が登場した⁵。LLM と汎用アームの組み合わせはシステムの構築・複製に必要な人・

金銭的成本を極限まで下げるので、ラボオートメーションの拡大に貢献できる。2023年はグラフニューラルネットワークのような特化型のAIと組み合わせた結晶探索とロボットの連携研究⁶が注目を集めた一方で、Natureレベルの論文³でもまだ室温・大気条件で進むシンプルな有機反応が報告される程度なので、本格的な化学・AI・ロボット連携は今後のテーマと言える。

大規模言語モデルによる化学推論と知識追加・ロボット制御・ポリマー合成

本研究では手始めに、化学分野におけるLLMの特徴と諸課題を整理することを目的に、GPT-4の能力をベンチマークで評価した^{4,7}。分子構造認識や化学反応のメカニズム解析などで大学院レベルの問題に適切に回答する能力を示し、反応条件の提案タスクではペイズ最適化を上回る性能^{3,4}も得られた。一方、化学への理解が不十分と思われる回答も多々観測された。次に、オープンソースの大規模言語モデルLlama 2 (Meta 2023)にオリジナルな化学文章やオープンアクセス論文を追加学習させる方法を検討した⁸。種々の制約があるものの、学習テキストに工夫を加えることで、LLMに専門知識を加えられることはわかった。

自家製のロボット装置や汎用アームを組み合わせた合成装置も構築中である。YOLO (You Only Look Once) という画像認識モデルを用いてガラス器具の座標を認識し、GPTと組み合わせることで、自然言語の指示に従って実験操作を行うシステムが得られている。本システムを用い、ポリイミド前駆体の合成などに挑戦している。



(左) ポリアミド酸の合成と熱イミド化のスキーム。

(中) プロトタイプの自動合成装置。

(右) ガラス器具を深層学習モデルで認識する様子。

1. W. X. Zhao, K. Zhou, *et al.*, 2023, **arXiv:2303.18223v11**.
2. LLM勉強会(LLM-jp)の想定する主な他分野展開先: 医療・法律・教育、マルチモーダル情報・ロボット制御 (<https://llm-jp.nii.ac.jp/about/>, 2023/12/21 閲覧)
3. D. A. Boiko, R. MacKnight, *et al.*, *Nature*, 2023, **624**, 570-578.
4. K. Hatakeyama-Sato, N. Yamane, *et al.*, *Sci. Technol. Adv. Mater.: Methods*, 2023, **3**, 2260300.
5. Google DeepMind RT-2: New model translates vision and language into action. <https://deepmind.google/discover/blog/rt-2-new-model-translates-vision-and-language-into-action/>
6. N. J. Szymanski, B. Rendy, *et al.*, *Nature*, 2023, **624**, 86-91.
7. K. Hatakeyama-Sato, S. Watanabe, *et al.*, *Digital Discovery*, 2023, **2**, 1548-1557.
8. K. Hatakeyama-Sato, Y. Igarashi, *et al.*, 2023, **arXiv:2312.03360**.

小規模データのマテリアルズインフォマティクスで挑む有機機能材料の探索と性能向上

(慶大理工) ○緒明 佑哉

Exploration and Enhanced Performances of Functional Organic Materials by Small-Data-Driven Materials Informatics (*Faculty of Science and Technology, Keio University*) ○Yuya Oaki

Our group has developed materials informatics combining sparse modeling and chemical insight for small data to optimize processes and explore new materials. I introduce the method of sparse modeling for small data (SpM-S) and its case studies to explore new functional organic materials for energy-related applications, such as battery and electrocatalyst.

Keywords : Small Data; Materials Informatics; Sparse Modeling; Conjugated Polymers; Energy-related Applications

マテリアルズインフォマティクス(MI)が注目されてから約10年が経ち、広く化学・材料への適用が進んでいる。MIを活用した有機機能材料の研究も盛んに行われている。近年では、ロボットやフロープロセスの利用、クローズドループの生成による自律自動化が急速に進んでいる。一方、有機機能材料創製へのMIの適用は、依然として様々な課題が存在する。例えば、データ量や質が十分でないことが多いこと、分子から材料に至る階層的なスケールの構造と機能の相関が複雑なこと、意外性と妥当性を両立した新規材料の提示と創出が難しいこと、研究者の熟練と経験が融合しにくいことなどが挙げられる。実験化学者である我々は、典型的な実験系のラボで得られる程度の小規模データをもとにプロセス最適化や物質探索を効率的に行うため、スパースモデリングと研究者の考察を融合した小規模データ向けMIを開拓してきた(Fig. 1)¹⁾。本発表では、その手法およびエネルギー関連材料を中心とした有機機能材料への適用事例について紹介する。

我々は、20–100個

程度の目的変数を含む小規模なデータを対象としてきた¹⁻¹²⁾。目的変数は、収率等の合成^{2,3,6)}、サイズ等の構造^{4,5)}、容量や反応電位等の機能⁷⁻¹¹⁾に関するパラメータとする。目的変数と相関があると考えられる説明変数を、研究者の経験や考察等をもとに挙げ、訓練データセットを作成する。続いて、機械学習を活用したスパースモデリングにより、寄与が大きな説明変数を記述子候補として抽出する。さらに、研究者の経

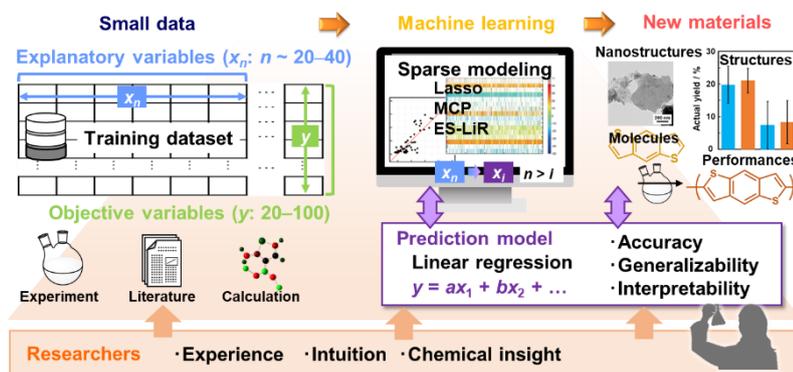


Fig. 1. Schematic illustration of our materials informatics combining machine learning and chemical insight for small data.¹⁾

験や考察等も含めて妥当な記述子を選定し、最終的に少数の本質的な記述子から構成される線形回帰の予測モデルを構築する。本手法では、説明変数の準備や記述子選定の際に、研究者の経験や考察を適度に融合することで、他の手法と比べて予測精度、汎化性(過学習の抑制)、解釈性を適度に両立した予測モデルの構築が可能となる^{1,12)}。得られた予測モデルを用い、検討したい条件や候補となる物質に対して予測に基づく仮想スクリーニングを行うことで、実験数を減らすことができる。

前述の手法、小規模データに対するスパースモデリング(Sparse modeling for small data (SpM-S))を利用し、リチウムイオン二次電池の有機正・負極活物質の探索および水電解による水素発生のためのメタルフリー電極触媒の探索を行った⁷⁻¹¹⁾。実験や文献から分子構造と反応電位・容量・エネルギー密度、水素発生過電圧などの性能値20~40個を目的変数として取得し、軌道準位や電解液との親和性などの20個前後の説明変数を準備した。SpM-Sによって正極の電位・容量・エネルギー密度⁹⁾、および負極の容量^{7,8,10)}、水素発生過電圧の予測モデルを構築した¹¹⁾。構築した予測モデルをもとに、候補となりうる低分子化合物の性能を予測した上で実験数を削減して実験を行った。その結果、いずれの系においても、既報と比較して優れた性能を示す新規化合物を見出すことができた (Fig. 2)。本手法は自前の小規模データに向いていることから、我々が進めている様々な新規の二次元・高分子機能材料への適用も期待できる¹³⁻¹⁶⁾。

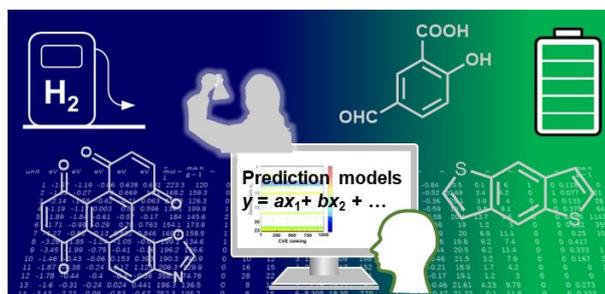


Fig. 2. Exploration of new anode active materials and metal-free electrocatalysts using SpM-S.^{7,8,11)}

- 1) Y. Oaki, Y. Igarashi, *Bull. Chem. Soc. Jpn.* **2021**, *94*, 2410 (Account). 2) G. Nakada, Y. Igarashi, H. Imai, Y. Oaki, *Adv. Theory Simul.* **2019**, *2*, 1800180. 3) K. Noda, Y. Igarashi, H. Imai, Y. Oaki, *Adv. Theory Simul.* **2020**, *3*, 2000084. 4) R. Mizuguchi, Y. Igarashi, H. Imai, Y. Oaki, *Nanoscale* **2021**, *13*, 3853. 5) Y. Haraguchi, Y. Igarashi, H. Imai, Y. Oaki, *Adv. Theory Simul.* **2021**, *4*, 2100158. 6) K. Noda, Y. Igarashi, H. Imai, Y. Oaki, *Chem. Commun.* **2021**, *57*, 5921. 7) H. Numazawa, Y. Igarashi, K. Sato, H. Imai, Y. Oaki, *Adv. Theory Simul.* **2019**, *2*, 1900130. 8) T. Komura, K. Sakano, Y. Igarashi, H. Numazawa, H. Imai, Y. Oaki, *ACS Appl. Energy Mater.* **2022**, *5*, 8990. 9) K. Sakano, Y. Igarashi, H. Imai, S. Miyakawa, T. Saito, Y. Takayanagi, K. Nishiyama, Y. Oaki, *ACS Appl. Energy Mater.* **2022**, *5*, 2074. 10) H. Tobita, Y. Namiuchi, T. Komura, H. Imai, K. Obinata, M. Okada, Y. Igarashi, Y. Oaki, *Energy Adv.* **2023**, *2*, 1014. 11) W. Hamada, M. Hishida, R. Sugiura, H. Tobita, H. Imai, Y. Igarashi, Y. Oaki, *J. Mater. Chem. A* **2024**, in press (DOI: 10.1039/D3TA06447F). 12) Y. Haraguchi, Y. Igarashi, H. Imai, Y. Oaki, *Digital Discovery* **2022**, *1*, 26. 13) J. Suzuki, A. Ishizone, K. Sato, H. Imai, Y. J. Tseng, C. H. Peng, Y. Oaki, *Chem. Sci.* **2020**, *11*, 7003. 14) M. Nakamitsu, K. Oyama, H. Imai, S. Fujii, Y. Oaki, *Adv. Mater.* **2021**, *33*, 200875. 15) N. Ono, R. Seishima, K. Okabayashi, H. Imai, S. Fujii, Y. Oaki, *Adv. Sci.* **2023**, *10*, 2206097. 16) N. Shioda, R. Kobayashi, S. Katsura, H. Imai, S. Fujii, Y. Oaki, *Mater. Horiz.* **2023**, *10*, 2237.

自動分子シミュレーションによる高分子物性データプラットフォームの産学共創

(統数研¹) ○林 慶浩¹

Co-creation of polymer properties data platform by automated molecular simulation through industry-academia collaboration (¹*The Institute of Statistical Mathematics, Research Organization of Information and Systems*) ○Yoshihiro Hayashi¹

In recent years, data-driven materials design techniques, known as materials informatics, have been rapidly introduced into the field of materials research. Needless to say, the most important resource of data-driven research is data. However, the amount of data on polymeric materials is overwhelmingly small, and at present there is practically no polymer property database that contributes to data-driven research. In order to develop a database of polymer properties based on molecular simulations, we have developed RadonPy,^{1,2} a Python library that supports the automation of polymer property calculations based on molecular dynamics (MD) simulations. RadonPy fully automates a series of processes required to perform MD calculations, including initial structure generation, charge and force field assignment, equilibration and nonequilibrium MD calculations, and calculation of physical properties. The vast computational resources of Fugaku and the industry-academia collaboration of 5 universities and 30 companies, led by the Institute of Statistical Mathematics, are producing data with the aim of constructing a database that includes more than 10^5 polymer skeletons.

Equilibration calculations for about 77,000 polymers were completed by the end of 2023 using RadonPy, of which thermal conductivity calculations for about 71,000 polymers and Tg calculations for about 40,000 polymers were completed. Comparison of the calculated and experimental values showed that there were large systematic biases and variations in specific heat and coefficient of linear expansion. The gap between the calculated and experimental values can be corrected by a machine learning method called transition learning, as shown in Fig. 1. To quantify the value of the database from simulations, we observed a scaling law for the prediction accuracy of transition learning with respect to

the number of simulated data. This scaling law has been shown to theoretically follow a power law, and we observed that the prediction accuracy improves in accordance with the power law for the RadonPy simulation data, as shown in Fig. 2. This scaling curve enables us to estimate the number of data required for the database and the achievable prediction performance, which is an important guideline for constructing the database.

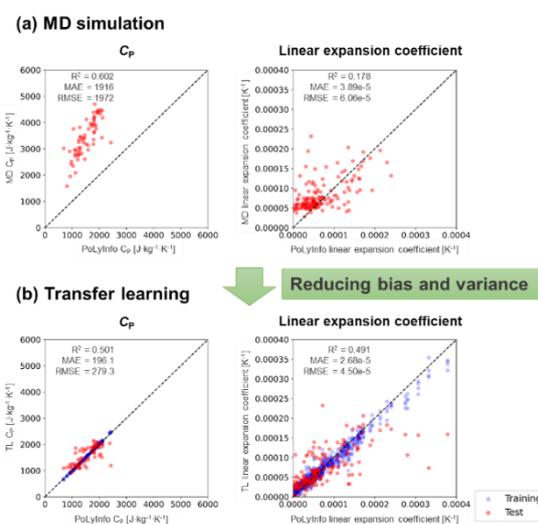


Figure 1. Calibration of MD calculated values by the transfer learning.

In addition, Bayesian optimization and automatic simulation using RadonPy are combined to design polymers with desired properties. An example of molecular design of polymers with both a high refractive index and a high

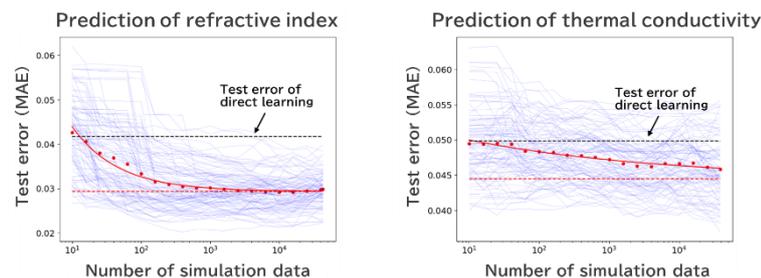


Figure 2. Scaling law of transfer learning using simulation data.

Abbe number, which are required for optical polymers, will be presented.

Keywords : Materials Informatics; Polymer Informatics; Molecular Dynamics Simulation; Transfer Learning; Industry-Academia Collaboration

近年、マテリアルズインフォマティクスと呼ばれるデータ駆動型材料設計の技術が材料研究の分野に急速に導入されている。データ駆動型研究の源泉は、言うまでもなくデータである。しかしながら、高分子材料のデータ量は圧倒的に少なく、データ駆動型研究に資するレベルの高分子物性データベースは現時点において実質的に存在しない。そこで、分子シミュレーションによる高分子物性データベースを開発するために、分子動力学 (MD) シミュレーションによる高分子物性計算の自動化を支援する Python ライブラリである RadonPy を開発した。^{1,2)} RadonPy は MD 計算の実行に必要な初期構造の生成、電荷・力場の割り当て、平衡化計算、非平衡 MD 計算、物性値算出の一連の工程を完全自動化する。そして、富岳の膨大な計算資源と、統計数理研究所が中心となり 5 大学・30 企業の産学連合体により、10⁵ 個を超える高分子骨格を包含するデータベースの構築を目指しデータ生産を行っている。

RadonPy を用いて 2023 年末までに約 77,000 ポリマーの平衡化計算が完了し、そのうち約 71,000 ポリマーの熱伝導率計算と約 40,000 ポリマーの T_g 計算が完了した。また、計算値と実験値を比較すると、比熱や線膨張係数において大きな系統バイアスやばらつきが存在した。この計算値と実験値の間のギャップは、Fig. 1 に示すように転移学習とよばれる機械学習手法により補正できることが示された。計算値によるデータベースの価値を定量的に示すべく、計算データ数に対する転移学習の予測精度のスケール則を観測した。このスケール則は、理論的にはべき乗則に従うことが示されており、RadonPy の計算データにおいてもべき乗則に従い予測精度が向上することが観測された。このスケール則カーブにより、必要なデータ数や到達可能な予測性能の見積もりが可能となり、データベースを構築するための重要な指針となる。

また、ベイズ最適化と RadonPy による自動シミュレーションを融合し、所望の特性を有する分子設計を行っている。当日は光学用高分子に要求される、高屈折率と高アッベ数を両立する高分子の分子設計の事例を紹介する。

- 1) Y. Hayashi, J. Shiomi, J. Morikawa, R. Yoshida, *npj Comput. Mater.* **2022**, 8, 222.
- 2) <https://github.com/RadonPy/RadonPy>

量子化学計算と機械学習を活用した有機 EL 材料の開発

(阪大院工¹) ○相澤 直矢¹Development of OLED materials with Inverted Singlet and Triplet Excited States (¹Graduate School of Engineering, Osaka University) ○Naoya Aizawa¹

Organic light-emitting diodes (OLEDs) are highly valued for thin, flexible displays but are limited by a maximum 25% efficiency due to the 1:3 singlet to triplet exciton ratio in electron-hole recombination. Researchers are addressing this limitation by using thermally activated delayed fluorescence (TADF) materials, which convert non-emissive triplet excitons into emissive singlet excitons through reverse intersystem crossing (RISC).¹ However, challenges such as slow RISC and efficiency roll-off at high currents persist. To screen TADF materials with fast RISC, this study combines density functional theory calculations and a machine learning technique, Bayesian optimization. The presentation will cover the virtual screening process, experimental validation, and an analysis of feature importance on fast RISC.

Keywords : OLED; TADF; DFT; Bayesian optimization; Virtual screening

次世代の有機 EL 材料として注目されている熱活性遅延蛍光 (TADF) 材料は、通常発光しない三重項励起状態から発光可能な一重項励起状態への遷移である逆項間交差により、有機 EL の内部量子効率を 100%まで高めることが可能である。しかし、逆項間交差が遅い場合、デバイスの劣化や高輝度時の発光効率の低下につながるため、より速い逆項間交差を示す TADF 材料の開発が求められている。量子化学計算や機械学習により、逆項間交差の速度定数 (k_{RISC}) の予測が可能となれば、材料開発のさらなる加速が期待できる。本研究では、量子化学計算とベイズ最適化を組み合わせた仮想スクリーニングにより、速い逆項間交差を示す TADF 材料の探索を行った。本発表では、仮想スクリーニングの結果や実験検証に加えて、速い逆項間交差に重要な特徴量について議論する (図 1)。

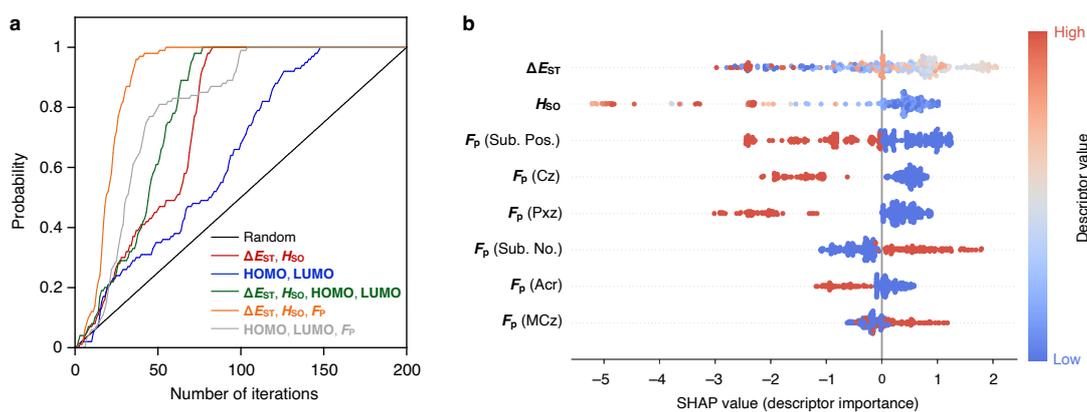


図 1. **a**, 様々な特徴量を用いたベイズ最適化の結果 (縦軸: 最大の k_{RISC} を見出す確率、横軸: 計算更新回数) **b**, SHAP (SHapley Additive exPlanations)による特徴量解析

1) H. Uoyama, K. Goushi, K. Shizu, H. Nomura, C. Adachi, *Nature* **2012**, 492, 234.

機能性ソフトマテリアルの構造・物性解明: 計算科学とデータ科学によるアプローチ

(北里大未来工¹・KISTEC²) ○渡辺 豪^{1,2}

Elucidation of Structure and Physical Properties of Functional Soft Materials: A Computational and Data Science Approach

(¹*School of Frontier Engineering, Kitasato University*, ²*KISTEC*) ○Go Watanabe^{1,2}

In general, functional soft materials are composed of organic molecules, so that they have high external field response and high flexibility. For these materials, computational science approach is one of the useful methods to understand the structure and physical properties between a single molecule and its molecular assembly. Data science is expected to play an important role in the development of novel high-performance functional materials. The presentation will highlight the new insights into the functional soft materials that have been using computational and data science methods.

Keywords : *Soft Materials; Molecular Dynamics Simulation; Data Science; Organic Semiconductors*

構成要素が有機分子であり、優れた応答特性とフレキシビリティを有する機能性ソフトマテリアルにおいて、実験のみでは分子スケールでの描像を理解することは容易ではない。そこで、原子や分子の動きを追うことができる分子動力学シミュレーションに代表される計算科学的手法に注目が集まっている。近年の CPU やメモリの飛躍的向上、GPU 等を活用した高速計算手法の確立により、様々な環境下に置かれた分子集合体の多彩なダイナミクスを明らかにすることが可能になってきている。

代表的な機能性ソフトマテリアルである有機半導体、特に低分子系は、原子種、置換基、側鎖の長さなどのわずかな分子構造の違いが集合構造、および物性に大きな影響を及ぼすことはよく知られている。この要因を解明するため、我々は分子動力学シミュレーションを用いて、結晶構造での原子レベルでの揺らぎ、いわゆる *dynamic disorder* の定量的解析に取り組んできた¹⁾。

また、革新的な有機半導体創製のためには、従来の実験主導の分子設計から脱却し、分子構造のみから結晶構造を始めとする集合体構造を予測する技術が不可欠である。最近、分子力学計算と分子動力学シミュレーションを組み合わせた手法で、n 型有機半導体の集合体構造予測が可能であることを見出している²⁾。さらに、機械学習を導入することで、より高精度で高速に集合体構造を提案する手法の開発にも着手している。

そして、有機エレクトロニクスデバイスの作製・評価においては、薄膜形成プロセスの影響を無視することはできない。これまでに、真空蒸着法で作製されるアモルファス有機蒸着薄膜で、膜形成過程を可視化し、ヘテロなマトリクス中における発光分子の配向性を評価し、分子構造との相関関係を議論している³⁾。

本講演では、上述のように、ソフトマテリアルを対象として計算科学やデータ科学によって得られた新しい知見を紹介する。

1) T. Okamoto, S. Kumagai, E. Fukuzaki, H. Ishii, G. Watanabe, N. Niitsu, T. Annaka, M. Yamagishi, Y. Tani, H. Sugiura, T. Watanabe, S. Watanabe, J. Takeya, *Sci. Adv.* **2020**, *6*, eaaz0632.

2) R. Ito, S. Sato, T. Seki, J. Takeya, T. Okamoto, and G. Watanabe, in preparation.

3) S. Shikita, G. Watanabe, D. Kanouchi, J. Saito, and T. Yasuda, *Adv. Photonics Res.* **2021**, *2*, 2100021.