

巻頭言「特集 マテリアルズインフォマティクスの最前線」について

吉田 亮[†] (オーガナイザー)

マテリアルズインフォマティクスは、材料科学とデータ科学の融合領域である。2011年に米国にてマテリアルズ・ゲノム・イニシアチブ(Materials Genome Initiative: MGI)という国家プロジェクトが始動した。材料開発には、新素材の発見から製品化までにおよそ10-20年という年月を要すると言われている。同プロジェクトでは、材料開発に要する期間を半分に短縮するという目標が掲げられ、そのホワイトペーパーにおいて材料データ基盤の整備とデータ科学の技術導入が目標実現への鍵になると宣言された。これを機にマテリアルズインフォマティクスという学際領域が一躍脚光を浴びることになった。我が国では、2015年にJSTイノベーションハブ構築支援事業「情報統合型物質・材料開発イニシアチブ」(拠点:国立研究開発法人物質・材料研究機構)が始動し、マテリアルズインフォマティクスの学術基盤の整備と人材創出、社会実装に向けた動きが急速に活発化した。統計数理研究所の研究者は、データ科学の学術基盤をもとにマテリアルズインフォマティクスの学術創生を促進してきた。さらに、2017年には統計数理研究所において「ものづくりデータ科学研究センター」が設立された[1]。同センターでは、素材・化学企業を中心に多数の企業との共同研究を推進し、材料開発の最前線で実践・実証研究を展開している。データ科学の独自の視点から材料研究の諸問題に対するユニークな切り口を発見し、新しい科学的手法を創出・実践する。これが同センターに課されたミッションである。本特集を構成する5報の論文は、この一連の取り組みから生み出された成果の一部をまとめたものである。

材料研究のパラメータ空間は極めて広大である。例えば、有機低分子化合物のケミカルスペースには、約 10^{60} 個の候補分子が存在すると言われている。さらに、実用材料の研究では、プロセスや添加剤・溶媒選択などが設計変数に加わり、パラメータ空間の大きさは爆発的に増大する。マテリアルズインフォマティクスの問題の多くは、このような広大な探索空間から所望の特性を有するパラメータを同定することに帰着する。これまでは実験と物理法則に基づく計算機実験が材料研究の進歩を牽引してきた。研究者の経験や勘に基づき材料を設計し、計算と実験による物性評価に基づき設計指針を見直す。このようなアプローチでこれまでに数多くの革新的材料が発見されてきた。しかしながら、経験や勘に基づく試行錯誤的な設計、計算、実験というループだけでは、決して超えられない壁が存在する。ここに「データ」と「データの科学」を導入することで、材料開発のコストを大幅に削減し、革新的特性を持つ新材料を創製する。これがマテリアルズインフォマティクスに寄せられた社会からの期待である。

2015年の「情報統合型物質・材料開発イニシアチブ」発足当時、マテリアルズインフォマティクスの世界は、ほぼ何もない広大なブルーオーシャンであった。当時この分野に集まった研究者には、データ駆動型科学が革新的な成果をもたらさしめる研究シーズを探索・発見し、自らのアイデアを実証し、社会に対してデモンストレーションすることが求められた。同セン

[†] 統計数理研究所: 〒190-8562 東京都立川市緑町10-3

ターの研究者らは、データ科学の学術基盤を活かし、この未踏領域を切り拓いてきた。こうした中で生み出されたいくつかの研究成果が本特集に取り上げられている。

吉田論文は、マテリアルズインフォマティクスの基本的なワークフローである順方向と逆方向の予測をデータ科学の視点から整理した。順問題の目的は、材料の構造から特性を予測するモデルを導くことである。これに対し、逆問題でモデルの逆写像を求めて所望の特性を有する材料を予測する。このワークフローを材料の“表現・学習・生成”という観点から整理し、データ科学の諸問題を論じている。また、材料研究における様々な逆問題を解説している。

吉田・ウ・森川論文は、高分子材料の設計における機械学習の適用事例を示している。同グループは、ものづくりデータ科学研究センターの研究者が中心となり開発したベイズ推論に基づく分子設計のアルゴリズム [2] を適用し、従来の高分子に比べて約 80% の高熱伝導率を有する新しい高分子の合成に成功した [3, 4]。本研究は、機械学習が自律的に設計した高分子が実際に合成された数少ない事例の一つである。本論文の趣旨の一つは、データ科学の実践の過程から浮かび上がってきたマテリアルズインフォマティクスが抱える問題点を切り取ることである。

ウ・山田・林・ザメンゴ論文は、高分子材料のマテリアルズインフォマティクスのレビューを行っている。マテリアルズインフォマティクスの研究対象の中でも、高分子という材料には独特の難しさがあり、研究の進展・実践展開に大幅な遅れが生じている。データベース、高分子の複雑性と表現の問題、材料特性の予測、設計という四つの切り口から、高分子インフォマティクスの現状と諸問題を論じている。

劉・山田・ウ論文は、材料研究の Small Data の問題を論じている。材料研究のデータの量は、データ科学の他の応用分野に比べると圧倒的に少ない。また、コミュニティ全体で Small Data を創出しようという動向も極めて低調である。Small Data の壁を乗り越えるデータ科学の技術である転移学習に着目し、複数の事例を紹介しながらその潜在的な学習能力を実証している [5, 6]。特に、転移学習を適用することで、データの範囲外に存在する材料の特性予測に成功した複数の事例を紹介していることが本論文の特色である。一般に革新的な材料の周辺にデータは存在しない。一方、従来のデータ科学では、基本的に入力が近ければ出力も近いという内挿的な予測を行う。転移学習から得られたモデルには、本論文で示された事例のように、外挿性が備わっていることがしばしば観測されている。

郭論文は、有機分子の合成経路の設計に関する機械学習の進展を解説している。コンピュータで合成経路を自動設計するという研究は、有機化学の分野で 50 年以上前から研究が進展してきた。その先駆者である有機化学者の Elias James Corey 博士は、2011 年にノーベル化学賞を受賞している。その後、脈々と継承されてきた研究の潮流が、近年の機械学習の進歩に合流することで、従来の発想とは全く異なる新しい技術を生み出そうとしている [7]。同論文を通じて、この転換期にデータ科学が果たした役割や技術変化の様子を垣間見ることができる。

本特集のタイトルは「マテリアルズインフォマティクスの最前線」である。最初に本特集を企画したのは 2019 年 3 月頃であった。この巻頭言を執筆している現在が 2021 年 6 月なので、本特集は 2 年以上前に企画されたことになる。マテリアルズインフォマティクスの学術体系はまだ発展途上の段階にある。このような技術の勃興が激しい創成期の学術領域において、2 年という年月はとても長く感じられる。本特集のトピックスは、企画当時は確かに「最前線」であった。しかしながら、2021 年 6 月現在、これらの研究はすでに最前線にはない。フロンティアの位置は時々刻々と変化している。本特集の著者らもすでに次のステージの研究に向かっている。5 年後に本特集を読み返したときに、我々はどう感じるのか。そんなことを本特集の著者らと話している。マテリアルズインフォマティクスはデータ科学、計算科学、実験科学の合流点に位置する。これらの全ての進歩が新しい科学的手法を生み出し、材料研究の在り方を刷新

していくに違いない。特に近年のデータ科学は、最先端の研究が応用分野に合流する時間差が急速に短くなってきている。5年後の「マテリアルズインフォマティクスの最前線」が果たしてどうなっているのか。退屈とは無縁の日々を過ごせることは間違いなさそうである。

1. 統計数理研究所プレスリリース (2017). 「ものづくりデータ科学研究センターの設立について」, <https://www.ism.ac.jp/noesuisin/news/monodukuri-opening.html>.
2. Ikebata, H., Hongo, K., Isomura, T., Maezono, R. and Yoshida, R. (2017). Bayesian molecular design with a chemical language model, *Journal Computer-Aided Molecular Design*, **31**, 379–391.
3. Wu, S., Kondo, Y., Kakimoto, M.-A., Yang, B., Yamada, H., Kuwajima, I., Lambard, G., Hongo, K., Xu, Y., Shiomi, J., Schick, C., Morikawa, J. and Yoshida, R. (2019). Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm, *npj Computational Materials*, **5**, 66.
4. 統計数理研究所プレスリリース (2019). 「機械学習の「記憶」を活用し、高分子の熱伝導性の大幅な向上に成功」, <https://www.ism.ac.jp/ura/press/ISM2019-07.html>.
5. Yamada, H., Liu, C., Wu, S., Koyama, Y., Ju, S., Shiomi, J., Morikawa, J. and Yoshida, R. (2019). Predicting materials properties with little data using shotgun transfer learning, *ACS Central Science*, **5**, 1717–1730.
6. 統計数理研究所プレスリリース (2019). 「物性予測タスク訓練済みモデルの包括的ライブラリ XenonPy.MDL を公開～転移学習で材料インフォマティクスのスモールデータの壁を乗り越える～」, <https://www.ism.ac.jp/ura/press/ISM2019-09.html>.
7. Guo, Z., Wu, S., Ohno, M. and Yoshida, R. (2020). Bayesian algorithm for retrosynthesis, *Journal Chemical Information and Modeling*, **60**, 4474–4486.