



いっぽう、上記の新聞記事からおおよそ10年を経た現在では、当初のMIへの期待が過熱していた状態を脱し、その有用性ととも材料開発への適用性も正しく認識されるようになってきていると思われる。

筆者は、偶然にもMIプロジェクトの立上げの場に立ち会い、産官学の研究者らとMIの可能性について議論する機会を持った。その当時、MIの秘めたる破壊力を実感しつつも、材料開発に実用化するためには多くの課題があると感じた。

MI関連の研究は日進月歩であり、限られた紙面で全貌を紹介することは容易ではない。今回、本稿を執筆する機会を得たので、菲才を顧みず、MI登場の背景から始めて、国内外の研究開発動向と研究事例、さらに、これからの材料開発はAI<sup>1)</sup>によってどのように変わるのかについて私見を紹介したい。

## 2. マテリアルズ・インフォマティクス (MI) 登場の背景—我が国の部素材産業を取り巻く環境—とMIの進展

我が国の部素材産業は、主要な輸出品目であるだけでなく、世界市場において高いシェアを有している。日系企業が世界シェアの60%以上を占

める製品のうち、約80%が部素材となっている。

そのなかでも機能性化学品は、日本企業が高い競争力を有する分野であり(図2)、たとえば、液晶ディスプレイの素材などの電子材料分野においては、日本の素材企業が高いシェアを有する。これはリチウムイオン電池の分野でも同様である<sup>5)</sup>。

一般的な素材開発は、扱う元素や反応条件の無限ともいえる組み合わせの中から目的に合う機能を持った素材を探し当てる作業である。以前は研究者の経験と勘が頼りで、なかには偶然の産物として開発された素材もある。社内に蓄積された経験やノウハウを利用し、競争力のある各種の素材を短時間で開発することが求められてきた。しかしすでに多くの素材が開発され実用化された現代では、より高性能で、要求される特性にマッチする素材を開発するのはコストがさらに嵩み、時間もかかるようになってきている。

いっぽう、我が国の部素材メーカーを取り巻く環境は、

i) 顧客ニーズが多様化し、バルクから少量多品種生産への転換、製品のライフサイクルの短期化など、研究開発のより一層の効率化や開発スピードのより一層の加速が強く求められるようになってきている。

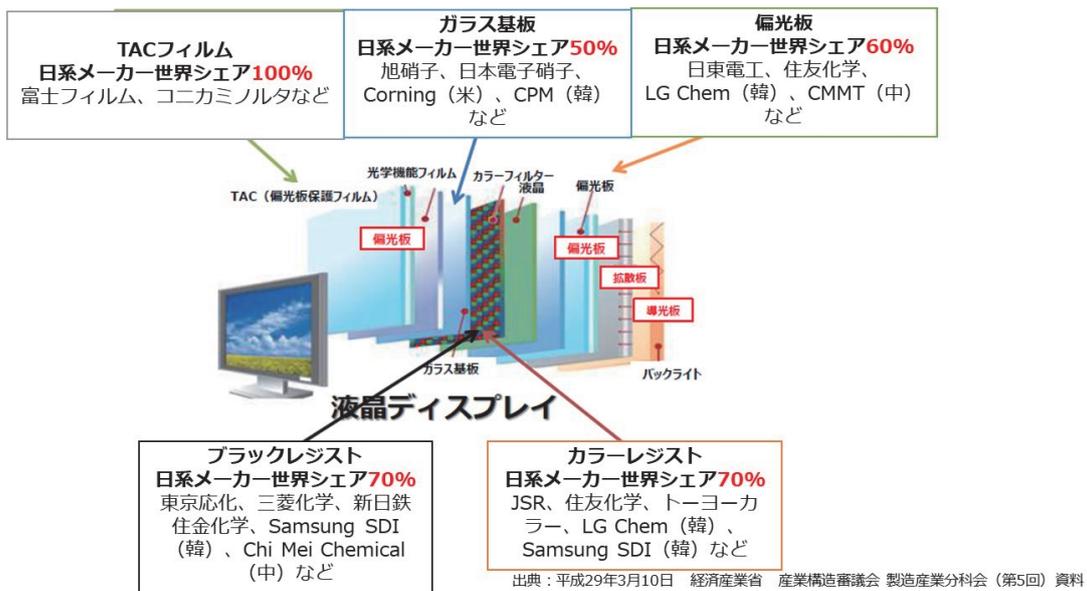


図2. 我が国の素材企業が競争力を有する機能性化学品（液晶ディスプレイ）の事例  
リチウムイオン電池も同様（吉野 彰博士（旭化成）ノーベル化学賞（2019））。

ii) グローバル化が一層進展し、海外顧客とすり合わせをし易い海外の部材メーカーの技術力が向上することによって、主要顧客であった国内顧客と、すり合わせを行いながら発展してきた、国内の部素材メーカーがシェアを失っている市場も出てきている、

iii) SDGs との関連で、働き方改革による生産性向上、社会の環境負荷低減への要求として、人体、環境への代替材料や蓄エネルギー／省エネルギー分野の研究開発が促進されているなど、大きな変革期にあるといえる。

我が国の部素材メーカーは、未だ技術的に優位性をもつものの、競争は激化してきている。新規材料の開発期間を短縮し、部素材メーカーの技術的優位性を維持・発展していくことが益々重要になってきている。

素材メーカーが現在、MI の導入を急ぐ大きな理由は、このような環境変化にある。問題解決には、「膨大な時間を必要とする材料開発の高速化やコスト削減」を狙いとした AI を活用するデジタル技術である MI の活用が不可欠で、MI の導入は世界規模で大きな流れとなっている。

### 3. MI とは何か

先に MI とはデータ科学と材料科学の融合領域であり、「データ駆動型材料開発」と言い換えることもできると述べた。ただ、これまでの材料科学の分野のデータは、それぞれの研究者のデータ形式がバラバラで交換ができなかったり、研究者ごとに抱え込んでいたり、同じ組成の物質でも求める特性が異なると、合成方法や蓄積されるデータが異なっていることなどの理由で、データ科学に必要なデータベースが構築されてこなかった。

しかし、物質の諸性質を電子状態に基づいて計算する第一原理計算により、コンピュータ上に材料がもつ基本的要素のデータベースを構築することが可能となった。これと経験に基づくデータを合わせて、そこに AI のような情報科学的手法を導入して未知材料の機能を推定できるようになった。さらに、この方法に多数の化合物群を一度に実験的に自動合成するコンビナトリアル合成や高速材料評価 (High Throughput Screening: THS)、論文からなどの知識抽出技術などを組合せることで、合理的かつ高速に所望の特性をもつ未知物質へと近づいていくこともできるようになった。これが MI における理想的なアプローチ方法である<sup>6)</sup>。MI による材料開発の流れは図 3 に示す概

#### マテリアルズ・インフォマティクス (MI) の概念図

実験データや計算シミュレーションデータ (データベース) などを用いた、機械学習 (AI) による所望の機能を実現する材料組成、構造の逆予測

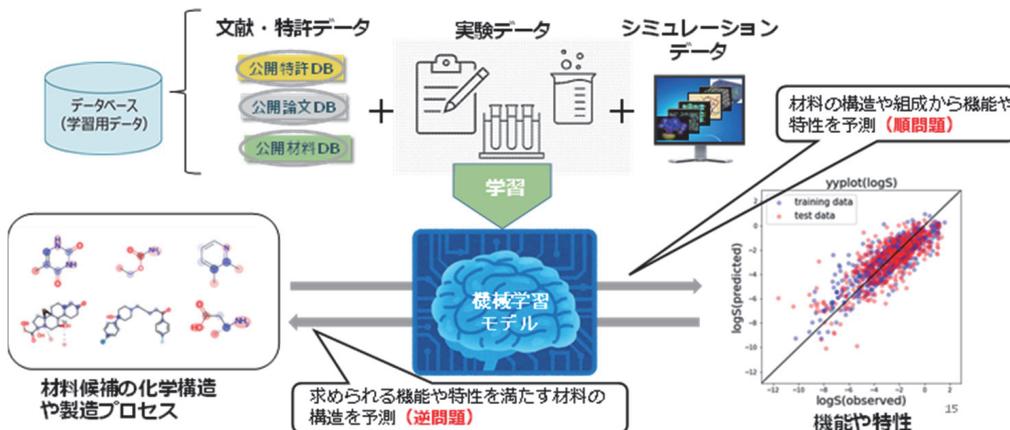


図 3. マテリアルズ・インフォマティクス (MI) の概念図  
実験データの部分にコンビナトリアル合成や THS を組み合わせられた、ロボットによる自動実験装置を導入することで、データ数を増やすことが期待できる。

念図で表される。

まず初めに、色々なソースからの材料とその機能に関するデータを大量に収集し、それらの相関関係を表すモデルを機械学習によって構築する。これを機械学習モデルという。これらデータは、化学構造と物性値の相関の実験データから数値シミュレーションデータ、科学論文・特許・社内文書といった文献情報までを包含するが、これら様々な情報をコンピュータが取り扱える形に数値データ化する必要がある（データの構造化）。

一度、適切な機械学習モデルができれば、このモデルに未知の材料の情報を入力すると、その材料が所望の機能をもつかどうか予測できる（順問題）。また、そうして得られたデータを用い、より精度の高い機械学習モデルを構築する。こうして得られた予測モデルに対して、目的とする機能・特性の情報を入力すると、その構造も予測できる（逆問題）。また、材料の代わりに材料の合成方法（プロセス条件）を用いれば、未知の合成法に対する性能予測にも利用できる。

図3からわかるように、企業や研究機関に蓄積された実験データは高速化を図るためのコアとなる部分である。

#### 4. 日本と世界の MI に対する取り組み

近年の材料・素材分野でのデータ活用の進展には、先述の米国でスタートした MGI を先駆けとする世界的な MI への取り組みが大きく影響している。MI は材料開発に情報科学手法を適用するアプローチであり、研究者や組織が蓄積してきたノウハウ等の暗黙知による材料の探索や改良を、大量の実験やシミュレーション結果等のデータ（テキストや図表、数値等の形式知）によって実行する。熟練者でなくても高度な開発ができる、開発期間の短縮につながるといった利点が期待される。

我が国でも MI に関連したプロジェクトが早い段階から立ち上がっている。文部科学省の新学術領域研究「ナノ構造情報のフロンティア開拓－材料科学の新展開」（2013－2017 年度）で実施され

た「材料科学と情報科学の調和」を嚆矢に、JST「情報統合型物質・材料開発イニシアティブ (MI<sup>2</sup>T)」（2015－2019 年度）、NEDO「超先端材料超高速開発基盤技術プロジェクト（超超プロジェクト）」（2016－2021 年度）、内閣府戦略的イノベーション創造プログラム（第2期）「統合型材料開発システムによるマテリアル革命」（2018－2022 年度）など、切れ目のない投資がなされ、材料科学と情報科学の研究者の連携推進や MI に関する研究拠点の整備が行われた。これにより材料データの蓄積や機械学習・データマイニングを用いた効率的な材料探索の手法、目的とする性能から構造・特性を提案し、それを実現するための最適な材料やプロセスを出力する逆問題 MI システムの開発など、今後の MI の普及拡大に必要なさまざまな取り組みが行われた。

2020 年 7 月に閣議決定の「統合イノベーション戦略 2020」において、マテリアルが戦略的に取り組むべき技術基盤として位置づけられ、データを基軸としたマテリアル DX プラットフォームの実現等が目標達成に向けた施策・対応策の 1 つとして示された。

また、同年 10 月から内閣官房において「マテリアル戦略有識者会議」が開催され、マテリアル戦略の 4 つの重要な項目の 1 つとして MI が取り上げられている。

そのいっぽうで、我が国全体として産学官のデータを効果的に収集・蓄積・流通・利活用する仕組みは整っていない点が今後の大きな課題とされた。世界を見渡した場合でも、上述の米国での MGI の下で各研究機関が材料研究データや文献等に関するデータプラットフォームを整備するなど一部で進みつつあるなど、データベースの構築は海外が先行しているが、産学官のマテリアルデータ全体を効果的に取り扱うための仕組み作りなどオープンデータの戦略的収集のための取り組みは各国模索中である。世界の主要データベースの分布を図4にまとめた。

このような背景から、材料分野におけるデータ

活用に向けたイノベーション政策として、2020年6月に文部科学省および経済産業省により同時発表された「マテリアル革新力強化のための政府戦略に向けて（戦略準備会合取りまとめ）」では「データを基軸としたマテリアル研究開発のプラットフォーム構築」が、今後当面推進すべき4つの取り組みの1つとして取り上げられており、国内の材料研究者が産学官の高品質な材料データを利活用できる環境（マテリアル DX<sup>7)</sup>プラットフォーム）の整備や材料の製造プロセスの高度化のための計測、シミュレーション技術、データ解析（AI等）を活用した基盤技術の確立等があげられている。

MIをめぐる各国の動きに関しては、紙幅の関係から表1にまとめた。

## 5. MI分野における研究動向

### 5.1. MIを用いた材料開発の動向

実際のMIによる材料開発ではAIの予測が百発百中するようなことはなく、機械学習だけで目的化合物の構造を予測することは、現時点では困難である。データ数が十分ではないため、精度の高い予測が得るのが難しいことが原因のひとつと考えられる。ただし、実際の材料探索では、多くの実験データが得られることは稀であり、限られたデータの中で探索を進めざるを得ない。

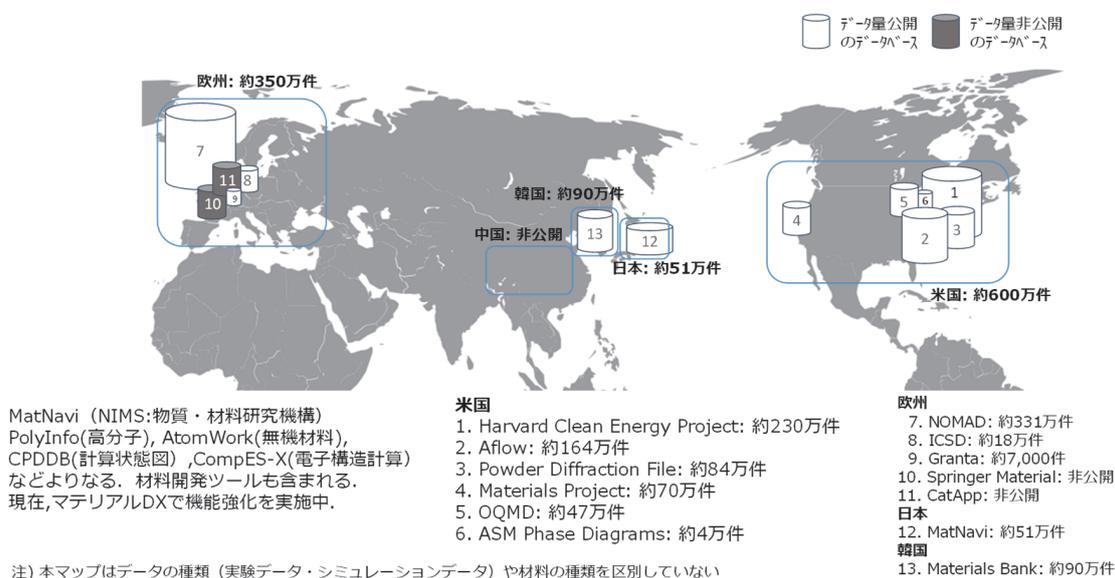


図4. 世界の主要データベースの分布図<sup>8)</sup>

- 

国立研究開発法人物質・材料研究機構 (NIMS) などが中心となり2013年以降、3つのプロジェクトが始動。全体予算は19年度までの累計で200億円規模。  
JST戦略的創造研究推進事業 CRESTやさきがけ (2015年頃～)  
NIMS 情報統合型物質材料開発イニシアティブ (MI2) 設置 (2015年～) 予算22億円  
AIST 超先端材料超高速開発基盤技術 (超超) プロジェクト設置 (2016年～) 予算138億円  
学会や雑誌でも多数の特集企画 (2019年頃～)  
「マテリアル革新力強化戦略」(内閣府 2021年4月～)において、「マテリアル・データと製造技術を活用したデータ駆動型研究開発の促進」(文科省 データ創出・活用型マテリアル研究開発プロジェクト事業 2021年～) 予算136億円(推定/10年間)
- 

11年に当時のオバマ大統領が国家戦略 “Materials Genome Initiative (MGI)”を発表し、世界で潮流が生まれるきっかけを作った。15年までの5年間で550億円以上を投資。
- 

14年からの7年間の科学技術計画 “Horizon2020” の枠組み内で、材料開発に特化した組織である、“Materials Modelling Council”を発足。
- 

米国の後を追うように中国版 “Materials Genome Initiative (MGI)” を立上げ、中央政府直轄の研究機関である中国科学院などが取組む。予算は100億円規模と見られている。『中国製造2025』も材料革命を支援。
- 

15年からの10年計画として “Creative Materials Discovery Project” を立上げ、300億円規模の予算が付けられた模様。

表1. MIをめぐる各国の動き

- 材料探索に「機械学習を使った場合」「そうでない場合」を比較すると、実験計画の回数が圧倒的に減少。
- 機械学習のアルゴリズムは、ベイズ推定をベースにしている。

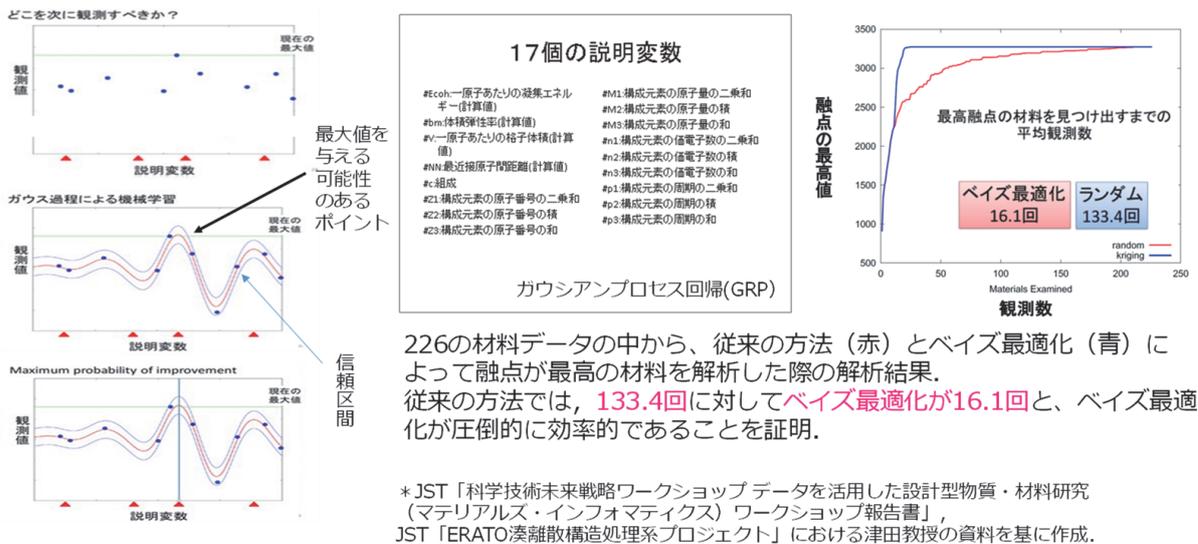


図 5. ベイズ最適化による二元化合物の融点データを用いた計算実験<sup>9)</sup>

したがって、代表的な応用事例は、観測済みのデータやデータベースからブラックボックス関数の式を推定するベイズ最適化、あるいはデータ量が少ない場合に用いられるスパースモデリングが主に用いられている。

## 5.2. 機械学習を用いたデータ駆動型材料設計技術の事例紹介

本項では、ベイズ最適化とスパースモデリングを用いた、それぞれの典型的な事例をとりあげ紹介する。

### 5.2.1. ベイズ最適化による二元化合物の融点データを用いた計算実験

あるポリマーを合成する際に、材料となるモノマー候補が13種類あるとすると、その組合せは理論上、100万通り以上存在する。そのため、いかに少ない実験条件で目的の組成を見いだせるかが課題になる。

こういった材料開発にMIの導入が期待される。ただし、多くの場合、データ数は多くはなく、高精度の予測モデルを構築することは現実的ではない。このような場合に有用なのが、ベイズ最適化という機械学習手法である。

二元化合物226個のなかから融点が最高のものを発見するという計算実験で、ベイズ最適化を用いた事例について紹介する。この事例では、17個の説明変数を用いて、i) 226個の材料のなかから5%をランダムに選んで融点を観測する、ii) その後、ベイズ最適化という機械学習手法を用いて、観測順を自動的に決定した(図5)。

この事例では、ベイズ最適化より得られた予測モデルを用いて、有望な実験ポイントが提案され、実際に実験して得られたデータを予測モデルにフィードバックするサイクルを回すアプローチとなっている。

ベイズ最適化では、獲得関数<sup>10)</sup>を設定することで探索(探索領域に見当をつける)と、活用(精度の高い領域での予測)のどちらに重点を置くかを選択できる。この事例では、最高融点の材料を見つけ出すまでの平均観測数は、ランダムに実施した場合は133.4回に対して、ベイズ最適化を用いた場合は16.1回となり、約8倍の効率化が達成されている。

### 5.2.2. スパースモデリングを利用して触媒反応の触媒活性(収率)を予測する技術の開発

触媒開発でもAIの活用が進められており、触

媒化学・人工知能・計算化学を融合し従来の触媒開発の方法を刷新する「キャタリストインフォマティクス」研究が報告されている。本法による触媒反応の収率や反応速度を予測する技術の開発について紹介する<sup>11,12)</sup>。

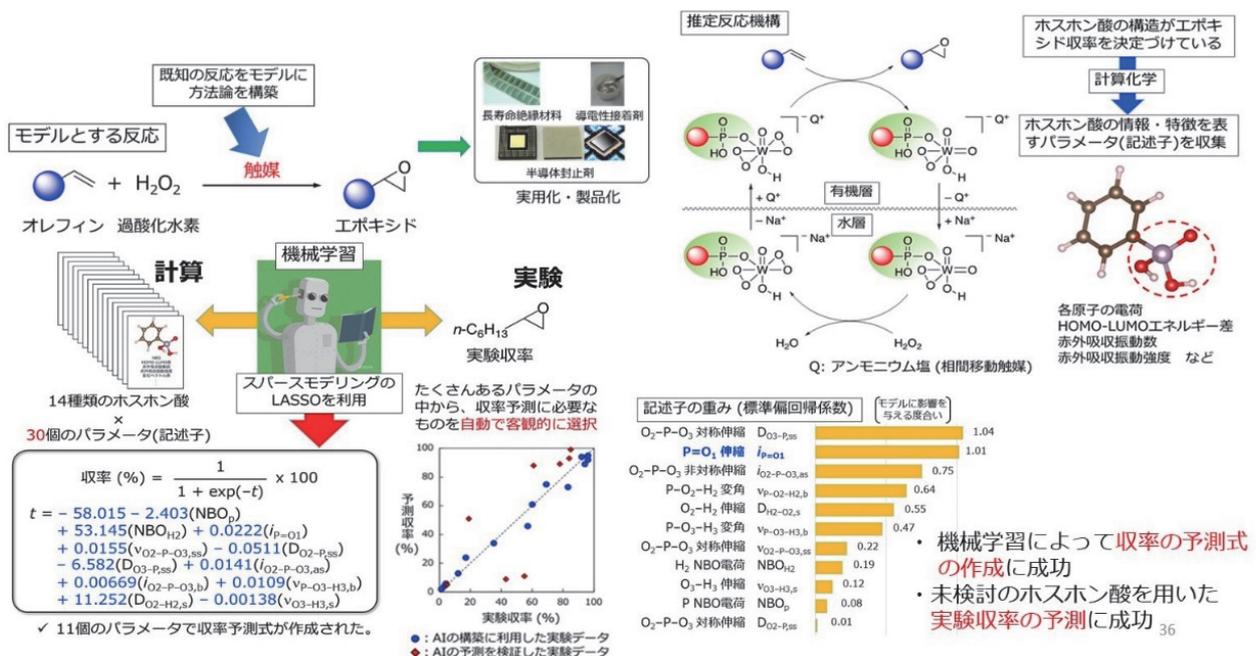
まず、過酸化水素を使ってオレフィンをハロゲンフリーでエポキシ化合物に変換する触媒反応をターゲット反応に選定し、機械学習を利用して本反応の収率を予測する技術の開発に取り組んだ。この反応はタングステン酸塩とアンモニウム塩とホスホン酸の3種の化合物を混合した触媒（三元系触媒）で進行する。各種ホスホン酸分子を検討した実験データに着目し、量子化学計算によってホスホン酸分子を数値化した原子の電荷や赤外吸収波数など30個のパラメータ（これを記述子という）を準備した。それらと実際のエポキシ化反応の実験収率とを相関づけて機械学習（スパースモデリング：代表的なアルゴリズムであるLASSOを使用）させ、収率予測モデルを構築することに成功した。この予測モデルにより、触媒活性を予測したいホスホン酸分子を用いたエポキシ化反応の収率が予測できることが明らかとなった（図6）。

この事例は、少数の実験データをもとに、スパースモデリングを用いて、収率の予測に寄与するパラメータを自動的に・客観的に選別して構築したAIで、収率が予測できることを最初に示した研究成果として知られている<sup>11)</sup>。30個のパラメータ（記述子）のうちに、機能やメカニズムに直接的に関わる記述子を含めていることがポイントである。この点に研究者としての経験やセンスが反映されている。

## 6. 今後の材料開発 – AIの導入で材料開発はどのように変わるのかー

MIの黎明期には、何でもAIで解決できるといった過度な期待が感じられた。しかし、現状のMIの技術水準では、材料開発に適用できる範囲は限定的であり、研究開発現場の課題解決に直接繋がらないケースもある。実際、実験科学、理論科学、計算科学（シミュレーション）側からアプローチの方が圧倒的に効率的である場合も多く、無理やりデータ科学側からアプローチすると、時間の浪費になってしまうこともある。

データ駆動型研究における最も重要な資源はデータである。しかしながら、現状では、材料研



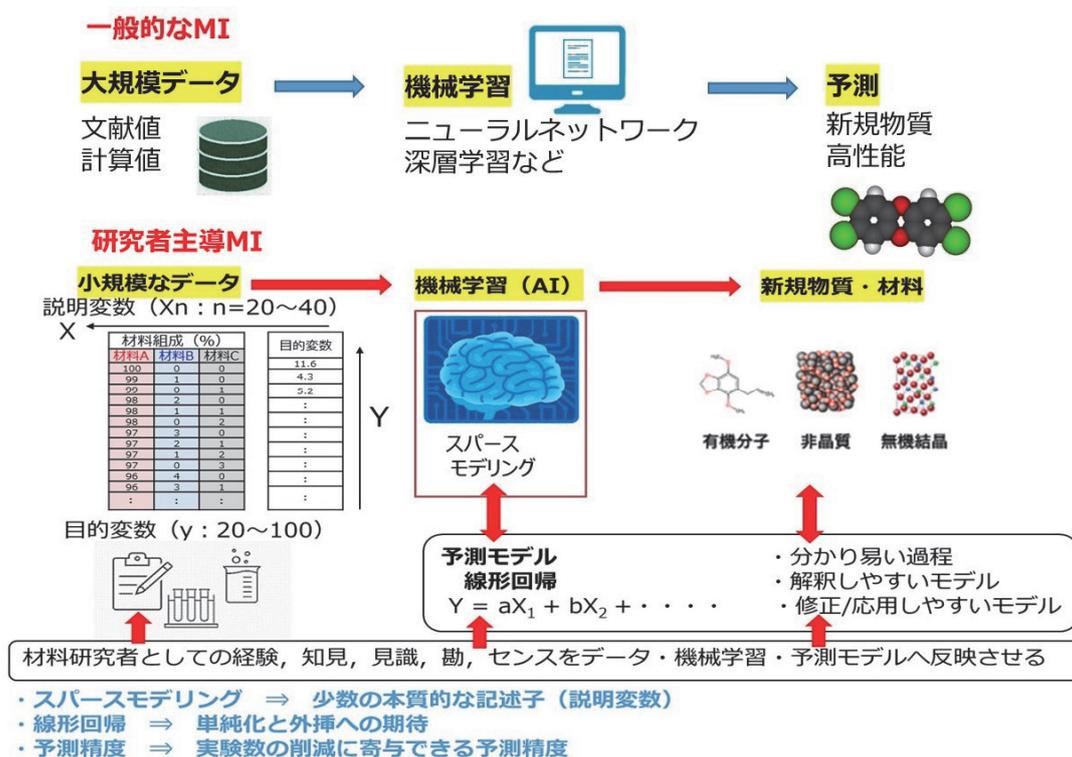


図7. 研究者主導 MI の進め方と一般的な MI との違い

究ではデータ駆動型研究に資するに十分なデータを利用できる段階には達していない。そのため、限られたデータの壁を乗り越えるため実践的な取り組みが盛んになってきている。

そのひとつが、5.2.1. で紹介した研究事例のような、研究者<sup>14)</sup>の経験と勘を重視した AI の利用である (図7)。これを「研究者主導 MI」と呼びたい。たとえば、研究者自前の小規模なデータ (精々100 個程度) と、経験と勘、研究者としてのセンスもフルに活用し、予測結果の解釈性も尊重しながら、大発見よりも、実験数削減などの効率化を目的に用いる方向に進展していくと思われる。

この場合、機械学習モデルはスパースモデリングにより構築されるが、予測モデルが線形回帰であれば、結果が単純でありブラックボックスにならないだけでなく、外挿の期待でき、研究者自身で予測結果の解釈ができる。したがって、機械学習モデルにもフィードバックでき、研究者が主導的に MI を使用できる。さらに情報科学者と緊密な連携・融合して、できれば、より良いシステムが構築できる。

かつて数値シミュレーションが期待と効用の乖離が大きかった黎明期を乗り越えて、今や必要不可欠なツールとなったように、今後、MI も研究開発現場では普通に利用されるようになるであろう。最近では、機械学習のライブラリが多数登場しており、Python, R, MATLAB など通じて、エンドユーザでも気軽に機械学習を取扱えるようになってきている。MI の活用ノウハウをいち早く蓄積するためにも、材料研究に MI を積極的に取り入れる環境も整ってきている。

今後、企業や研究機関で高品質な材料データ基盤が整備され、目的変数と複数の説明変数の関連性を見出すための大量なデータセットが取り扱えるようになれば、機械学習や統計学を活用する幅が広がる。

また、実験データの部分に、コンビナトリアル合成や THS を組み合わせたロボットによる自動実験装置が導入できれば、高品質な材料データが短時間で得られるようになる。MI の今後の発展を考える際には、解析技術以外の様々な技術動向も注視していく必要がある。

## 文献および脚注

- 1) S P Ong, Y Mo, W D Richards, L Miara, H S Lee, and G Ceder, *Energy & Environmental Science* 2012, 6 (No.1) 148
- 2) Kamaya, N.; Homma, K.; Yamakawa, Y.; Hirayama, M.; Kanno, R.; Yonemura, M.; Kamiyama, T.; Kato, Y.; Hama, S.; Kawamoto, K.; Mitsui, A. *Nature Materials* 2011, 10, 682.
- 3) この記事自体には幾つかの事実誤認が認められる。主なものは次の3点である； i) 記事に掲載の材料は、全固体リチウムイオン電池の固体電解質材料として用いる結晶性の硫化物  $\text{Li}_{10}\text{GeP}_2\text{S}_{12}$  (LGPS) であって「電極材料」ではない、ということ、 ii) LGPS を発見したのはトヨタ自動車、東京工業大学の菅野了次教授、高エネルギー加速器研究機構 (KEK) の研究チームであって日本企業 (単独) ではない、ということ、 また、 iii) 米韓チームは、日本のグループが *Nature Materials* 誌\* で発表した LGPS について、Ge と S を他の元素で置換した場合における結晶構造の安定性、電気化学的安定性、リチウムイオン導電性をシミュレーションで評価している。米韓の研究者は、日本のグループの発表した論文の組成や結晶構造の情報に基づき、その改変材料を検討したのが実態であって、ゼロベースから導き出したのではない、の3点である。
- 4) 広義に範囲を持つ AI 分野のひとつの領域が機械学習であり、本稿における AI とは機械学習のことを意味している。
- 5) 経済産業省 産業構造審議会 製造産業分科会 (第5回) 資料 (平成 29 年 3 月 10 日)
- 6) 田中功, 世古敦人, 森分博紀, *FC Report* 2016, 34 (No1), 8.
- 7) DX とは、デジタルトランスフォーメーションの略称である。
- 8) 国立研究開発法人新エネルギー・産業技術総合開発機構 平成 28 年度成果報告書情報収集事業 マテリアルズ・インフォマティクス等に関する周辺動向調査 (2017 年 04 月 15 日) の 71 頁の図を基に筆者作成。
- 9) JST 「科学技術未来戦略ワークショップ データを活用した設計型物質・材料研究 (マテリアルズ・インフォマティクス) ワークショップ報告書」
- 10) 獲得関数とはベイズ最適化において得られた事後分布から次の候補点を探索するための関数のこと。獲得関数はガウス過程回帰により得られた事後分布から候補点を選択する際に用いる。ガウス過程回帰からグラフのように平均値と分散が表される。獲得関数はこの平均値と分散から目的に応じて候補点を選択する。
- 11) Yada, A.; Nagata, K.; Ando, Y.; Matsumura, T.; Ichinoseki, S.; Sato, K. *Chem. Lett.* 2018, 47, 284.
- 12) Yada, A.; Matsumura, T.; Ando, Y.; Nagata, K.; Ichinoseki, S.; Sato, K. *Synlett*, 2020, eFirst (DOI:10.1055/a-1304-4878) .
- 13) 矢田陽, キャタリストインフォマティクスの現状と将来への期待, H30 年度第 2 回特別フォーラム「最近の化学におけるインフォマティクスの動向」(2019)
- 14) ここでいう研究者とは、材料科学, そのなかでも合成領域を専門とする研究者を指す,