

# 第 1 章

## マテリアルズ・ インフォマティクスとは

技術的な説明に入る前に、まずはマテリアルズ・インフォマティクスの全体像を大雑把におさえましょう。



## 1.1 機械学習とマテリアルズ・インフォマティクス

近年、様々な分野で人工知能（AI：Artificial Intelligence）の技術に期待が寄せられています。皆さんに大きなインパクトを与えた出来事といえば、AIによるボードゲームでしょう。例えば、2016年に囲碁のAI（AlphaGo）が世界最強のプロ棋士（イ・セドル氏）に勝利し、その翌年、将棋のAI（Ponanza）は叡王戦覇者の佐藤天彦名人を破りました。AIが、その領域のトップに君臨する方々を打ち負かしてしまったことは、世界中に衝撃を与えたと同時に、人々に期待と不安をもたらしました。

そのAIの中核をなすものが、機械学習と呼ばれる統計数理に基づいた技術。そして、この機械学習を含む情報処理技術をフルに活用し材料開発を進めていく分野を、マテリアルズ・インフォマティクス（MI：Materials Informatics）と呼びます。MIは、この機械学習に加え、物性理論、実験、シミュレーション、データベース、クラウド、ロボティクス、プラットフォーム、IoT、通信、セキュリティ……などなど、様々な技術が統合されて成り立っている分野といえます。本書ではその中でも機械学習に焦点を当て、その概要を実際の材料開発への応用事例を踏まえて記載していきます。また、MIに関わる他の情報や技術は、コラムにて記載いたしますので、そちらも参考にしてください。

## 1.2 機械学習さえあればすべて解決するの？

さて、MIの非常に大雑把なイメージが図1.1に描かれています。科学者と人工知能が一緒になって材料開発をしていくというイメージです。この図で重要なのが、MIの主役は物理・化学・材料学などの専門知識を持った科学者、つまり人間であり、人工知能や機械学習は、あくまで人間をアシスト（補助）するツールであるという点です。

囲碁AIや将棋AIの場合は、すでに人間を凌駕した存在になってしまったと

## 1.2 機械学習さえあればすべて解決するの？

言えますが、それは、ボードゲームで解くべき問題が比較的単純だからです。例えば将棋の場合、試合の進め方や駒の動かし方といった“ルール”が明確に決められており、相手に勝つ（相手棋士の王将をとる）という“目的”もはっきりしています。こういった問題はAI（機械学習）に任せやすい領域です。一方、材料開発の場合は特に決まった“ルール”というものがありません。どんな材料をどのように作るかは自由です。このように“ルール”が曖昧な分野の問題を、機械学習で解くのは大変です。例えば、将棋AIに将棋のルールを与えず、相手に勝つ（相手棋士の王将をとる）という目的だけ与えたとしましょう。するとどうなるでしょうか。もしかしたら、『相手棋士を殴って王将を奪い取る』という行動が、目的を達成するために最も効率的な選択になるかもしれません。将棋のルールが与えられている場合は、盤面のことだけを考えていけばよいのですが、ルールが曖昧になった途端、相手棋士の筋肉量・反射神経など色々なことを考慮する必要が出てきます。上記は極端な例ですが、こ

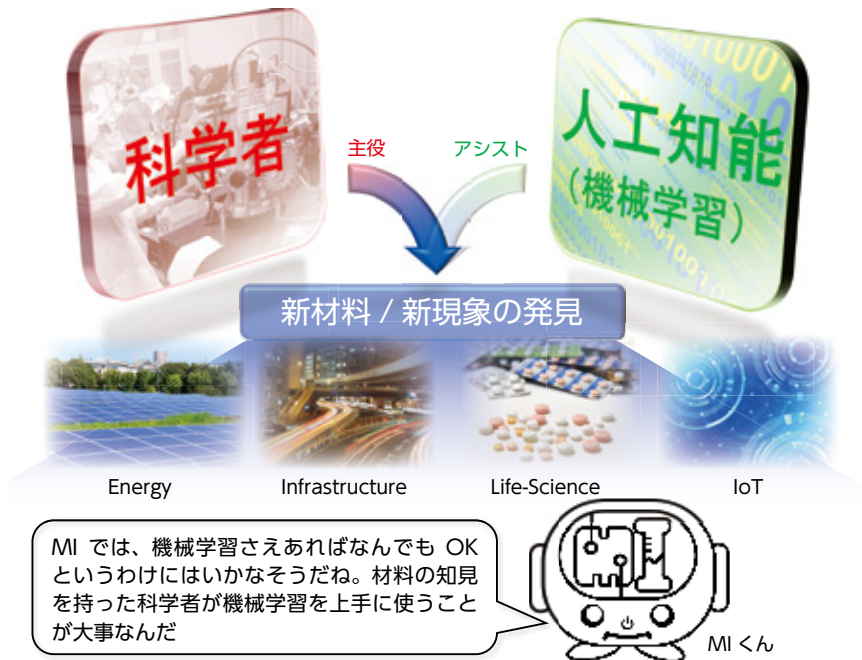


図 1.1 マテリアルズ・インフォマティクス

のようにルールが曖昧になると、考えなければならない事柄が爆発的に増え、問題が難しくなります。

材料開発は、まさにこのルールが曖昧な分野です。そのため、AIや機械学習さえあればすべて解決する分野ではありません。人間が、物理・化学・材料学などの専門知識を使って上手に材料開発における問題設計（ルールや目的の設定）をし、その問題を解くのに適した機械学習を使いこなしながら、材料開発を進めるということが重要になるのです。

## 1.3 理論科学・計算科学・実験科学・データ科学（機械学習）

MIによって、機械学習という技術が材料開発の分野に導入されたわけですが、これによって従来の材料開発の技術（理論科学・実験科学・計算科学）が不要になったわけではありません。役割が全く異なります。図1.2にそれら4技術の関係性を示します。これらの技術は、それを主に実行する対象（人間 or コンピュータ）と、アプローチの種類（演繹的 or 帰納的）で分けることができます。演繹的とは、論理的形式に頼って事実（ファクト）を1つずつ積み重ねていくアプローチを言います。材料開発の分野において、演繹的アプローチを人間が実空間で実行していくことが理論科学です。そして、この演繹的アプローチをコンピュータによってサイバー空間で行っていくことが計算科学、つまりシミュレーションに当たります。一方、実験科学は帰納的なアプローチです。実験によってある材料を作成し、その材料特性を測定（評価）したということは、いわゆるその材料に関する答えのような情報を得たこととなります。我々は、その情報をもとに、「なぜこの材料でこの特性が得られたのか」ということを逆説的に考えていきます。そのため実験科学は、人間が実空間で帰納的アプローチによって材料開発をしていることに相当します。データ科学（機械学習）は、この帰納的アプローチをコンピュータ内（サイバー空間）で行っている領域に分類されます。

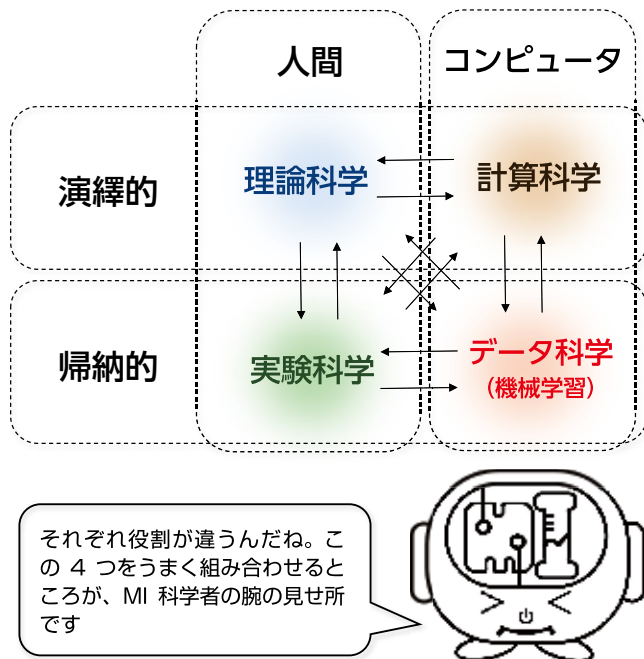


図 1.2 理論科学と計算科学と実験科学とデータ科学

## 1.4 演繹的と帰納的

この、演繹的と帰納的の部分についてももう少し考えてみます。図 1.3 の左下には材料の構造 (Structure) や組成 (Composition) の情報があり、右上に材料の機能 (Function) や特性 (Property) の情報があるとしましょう。この図において、構造・組成の情報から機能・特性の情報を導き出すアプローチを、MI 分野では演繹的と表現したり、または“順問題を解く”と表現したりすることがあります。このアプローチは、材料 (物性) の基礎研究をやっている大学の先生がよく採用するアプローチです。例えば、カーボンナノチューブ (CNT) という新しい物質が突然見つかったときのことを考えましょう。人々は、このカーボンナノチューブという物質について「どういった材料特性があ

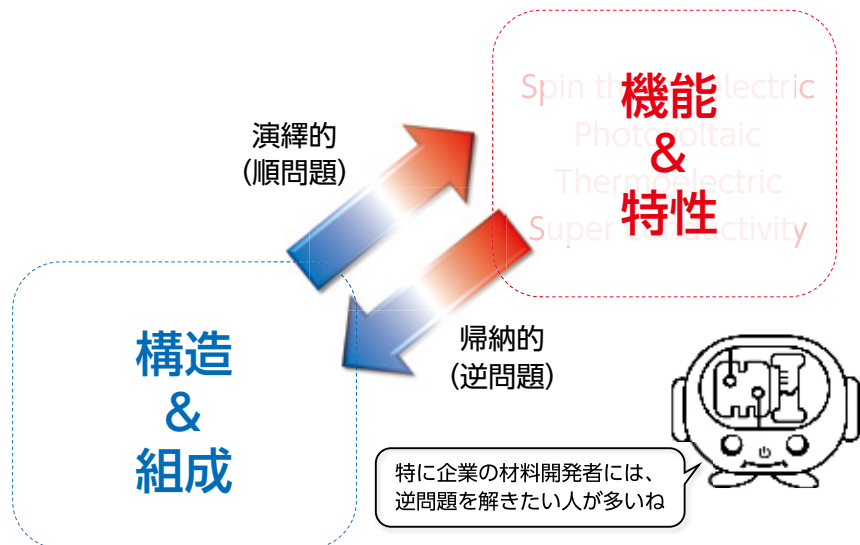


図 1.3 順問題と逆問題

るのだろうか」「何かに応用できないだろうか」ということを考えながら材料開発を進めてきました。つまり、カーボンナノチューブという与えられた構造・組成の情報から、この物質にある機能・特性の情報を明らかにしようと頑張ってきたわけです。まさに図1.3の順問題を解こうとしています。これに対して、もう少し応用寄りの材料開発者（特に企業の材料開発者）は、全く逆方向のアプローチをとることが多いです。例えば、熱電材料を開発する状況を考えてみます。この場合、「熱電効率が良い材料を開発したい」であったり「安価な元素で熱電材料を作りたい」であったりと、明確な目的が存在します。そして、材料開発者はその目標に向かって、「どんな構造・組成がよいのか」ということを考えながら開発を進めます。これはつまり、目的（材料の機能や特性）を達成するために、材料の構造・組成を探していることになります。MIの領域ではこういったアプローチを帰納的、または“逆問題を解く”と表現します。帰納的なアプローチをとれるデータ科学（機械学習）の技術が材料開発に導入されたことによって、この材料の逆問題を解くことが昔に比べてやりやすくなりました。