

マテリアルズ・インフォマティクスに基づいた新材料の探索

京都大学工学研究科 材料工学専攻

田中 功, 世古 敦人, 大場 史康, 東後 篤史

無機化合物の多様性と材料探索

材料科学で取り扱う元素の種類は 80 程度であるが, その組み合わせは 2 元系で約 3 千, 3 元系になると約 8 万となる. しかし可能な化合物の数は, 組成比を考慮すると大きく増える. 化合物 $A_xB_yC_z$ で x, y, z を整数として, $x + y + z = 10$ と限定しても, その組み合わせは 100 とおりであり, 3 元系全体では 800 万とおりになる. 同じ組成であっても結晶多系と呼ばれる様々な構造を持つ場合が多いので, その数はさらに増える. 一方で, 世界最大の無機結晶構造データベース(Inorganic Crystal Structure Database: ICSD) に収録されている 3 元系までの無機化合物データは現在 76000 件であり, 同一の化合物で複数のデータが収録されている重複を除けば, 実験的に結晶構造が既知の無機化合物の数は 5 万件を下回る. このうち生成エネルギーなどの熱力学データが既知である化合物は数千件であり, 物性値が既知である化合物の数はさらに少ない.

材料探索とは, 一般に既知物質の既知特性に対し, 別の物質で同等あるいは, それを越える特性を実現することを指す. 別の物質とは, 構造未知の新物質だけでなく, よく知られた物質であっても, 当該特性について注目されていない場合を含む. 既存の材料の特性を向上させるために構成元素や組成を探索する研究は広く行われている. 高価な希少元素を安価で豊富な元素で代替して同一の性能を実現するというのも, 材料探索の重要な目的のひとつである.

経験則による物質の整理と探索

材料探索は, 材料研究の主要テーマであり, ひとりの研究者が勘と経験に基づいて多くの物質を合成して特性評価するということが広く行われてきた. その中で, 構造や特性などの実験データを様々な因子で「整理」し, その相関性をもとに機能発現のメカニズムを推定, そして理論モデルを構築するという手法は材料探索に大きく貢献した. 典型例がライナス・ポーリングの提案した諸概念である. ポーリング以来, 現在に至るまで 80 年近く電気陰性度やイオン半径といった単純な説明変数を用いた様々な経験的法則に基づいて材料探索が行われてきた. その結果, 多くの新規材料や機能が「発見」されてきたことは事実であるが, このような経験則による探索に限界があることも認識されてきた. ポーリング則を超越する (beyond Pauling) というのは, ポーリング則に従わない物質を探索するという意味ではなく, 単純な経験則に依ることなしに材料探索を行うという意味での象徴的な標語である.

マテリアルズ・インフォマティクスの提案

高速、大容量化した計算機を利用して膨大な情報を統合・整理したうえで、必要な知識を取り出すデータマイニングと呼ばれる技法が、自然科学・社会科学の多様な分野で活用されている。自然科学への応用の代表例がバイオインフォマティクスと呼ばれる生命科学と情報科学との融合領域である。これは遺伝子に関する膨大な実験データを計算機で解析し、その情報に基づいて生命現象を理解するというもので、すでに大きな成果を上げている。同様に、計算機の巨大情報処理能力を材料研究に応用する学術領域を私たちは**マテリアルズ・インフォマティクス（材料インフォマティクス）**と呼んでいる。材料に関する構造や特性など様々な情報をデータベースとして集約し、それを適切に整理することで、材料探索に利用するのである。

高精度第一原理計算に基づいたマテリアルズ・インフォマティクス

データマイニングを行うためには、十分な情報を収録したデータベースが必要である。しかし、先述したように、対象とする材料の多様性に対し、既知のデータ数はあまりに少ない。

最近、量子力学の原理のみに基づいた電子論計算、いわゆる第一原理計算に長足の進歩があった。計算機の性能と効率的な計算手法の出現、さらに計算精度が大幅に向上したことで、結晶構造、電子（磁気）構造、フォノン状態、生成自由エネルギー、誘電率、弾性率などの情報を温度や圧力の関数として定量予測することが可能になってきた。このような物質の一次情報についての計算結果をデータベース化すると、実験結果のデータベースに比べて、情報量が飛躍的に増大するとともに、情報を均質に揃えることが可能となる。

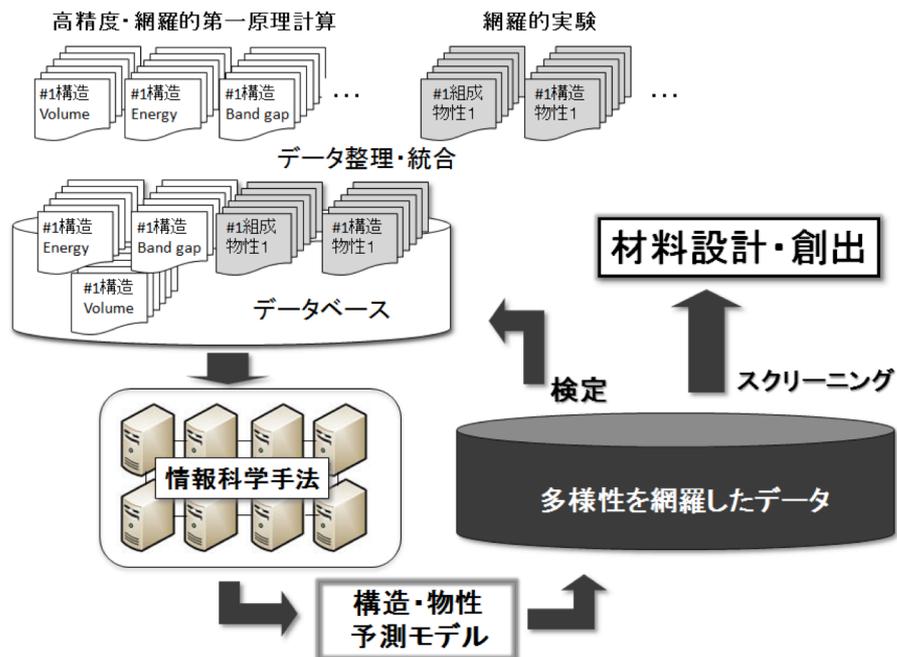
しかし、このようなデータベースを構築するだけでは、未だ材料探索として不十分である。主に2つの理由を挙げることができる。第一の問題は、多元系における化合物の巨大な自由度である。先述したように、単純組成の3元系物質だけで800万とおりである。さらに多元系を考えると、全ての物質を網羅することが現実的でないことが理解できる。さらに多くの物性が化学組成に対して加算的ではなく、多元系における物性予測は単純作業ではない。第二の問題は、多くの材料機能が、物質の一次情報から直接的に演繹されるものではない点である。多くの材料機能は、様々なスケールでの階層構造の性質が創発的に現れたものであり、物質の一次情報と材料機能の関係は自明ではない。さらに材料機能の多くは、基底状態についての静的情報だけでなく、物質内のイオン、電子、熱のトランスポート機能に伴う励起状態についての動的情報を必要とする。

このような問題点を認識した上で、現実的に獲得可能な数の第一原理計算結果を元に、未知である材料機能を合理的に推定し、それに基づいて物質のハイスループット・スクリーニングを行うというのが、私たちが提案している第一原理計算に基づいたマテリアルズ・インフォマティクスである。

材料機能の合理的推定と検証

材料機能の推定は、化合物の材料機能のデータから何らかの説明変数をもとに目的変数を獲得するデータマイニングと位置づけられる。具体的な方法としては、化合物の材料機能の定量的データから、予測化合物の材料機能を定量的に獲得する回帰分析、あるいは材料機能の優劣に従ってグループ分けされた化合物のデータをもとに、予測化合物の属するグループなどの離散的な材料機能を獲得する判別分析が挙げられる。すでに情報科学においてロジスティック回帰、ニューラルネットワーク、サポートベクターマシンなど多様な解析手法が提示され、広範に利用されている。第一原理計算に基づいた材料機能の推定を行う場合には、説明変数として、温度、圧力、化合物の組成などの外部条件や、原子番号やイオン化エネルギー、原子量のような元素についての根源的な値のほか、第一原理計算によって獲得した物質の一次情報を利用することが考えられる。

そして、得られた予測の精度を上げるためには、予測モデルの検証を逐次行うことが有効である。すなわち、予測された結果について、予測に利用しなかった独立データを用いた検定を行うのである。予測モデルの妥当性は、この独立データを用いた検定時の誤差によって評価することができる。材料探索においては、最終的に実験を行うことが想定されるので、検定のための独立データとして、実験値を検定に用いることが自然である。しかし第一原理計算によって材料機能をシミュレーションすることが可能である場合には、予測モデルの構築段階で計算結果を利用することも考えられる。たとえば第一原理分子動力学計算に基づいて、複雑結晶における原子やイオンの拡散現象をシミュレーションする場合などがそれに対応する。



高精度第一原理計算に基づいたマテリアルズ・インフォマティクスの模式図