

## マテリアルズ・インフォマティクス：機能無機材料探索における機会と課題

### Materials Informatics: Opportunities and Challenges for Inorganic Materials

物質・材料研究機構 MaDIS<sup>1</sup>, 総合科学研究機構<sup>2</sup>, アルバータ大学<sup>3</sup>, 北海道大学<sup>4</sup>

○田中譲<sup>1-3</sup>, 藤間淳<sup>1</sup>, 高橋啓<sup>1,4</sup>

NIMS MaDIS<sup>1</sup>, CROSS<sup>2</sup>, University of Alberta<sup>3</sup>, Hokkaido University<sup>4</sup>

Yuzuru Tanaka<sup>1-3</sup>, Jun Fujima<sup>1</sup>, Keisuke Takahashi<sup>1,4</sup>

[TANAKA.Yuzuru@nims.go.jp](mailto:TANAKA.Yuzuru@nims.go.jp), [FUJIMA.Jun@nims.go.jp](mailto:FUJIMA.Jun@nims.go.jp), [keisuke.takahashi@eng.hokudai.ac.jp](mailto:keisuke.takahashi@eng.hokudai.ac.jp)

先端科学技術分野では、種々の‘X’サイエンスが、‘X’インフォマティクスへとパラダイム・シフトを始めている。最初の変革はバイオ・サイエンスで起こり、バイオメディカル・サイエンスが続いた。このパラダイム・シフトは今後あらゆる分野で進み、計測・観測機器や生産プロセスはセンサとアクチュエータを用いて完全にコンピュータ制御され、ハイスループットの実験やシミュレーションの設定とデータ取得、仮説設定、データ分析による仮説検証、文献から抽出された知識ベースへの問い合わせ、報告書作成までのすべてがコンピュータ端末を通じて行われるイン・シリコ・サイエンス (in silico science) の時代の到来が目指されている。Xインフォマティクスの発展は4世代に分かれる。第1世代では、データ駆動型研究への期待が高まり、コンピュータ・サイエンス (CS) の専門家が機械学習手法を導入して、領域専門家を助けて初期の成功を挙げる。やがて機械学習の単純適用による課題解決は限界を迎え、領域知識を駆使して情報学的/数学的に対象の構造や機構を捉えなおし、機械学習に適した数理モデル構築の必要性が認識され、ブレークスルーの模索の時代である第2世代が始まる。機械学習に適した対象の新しい数理モデリング、質の高いデータ取得のためのハイスループット実験とシミュレーション、機械学習に適した対象の人工的なデザイン・アーキテクチャ等のブレークスルーの発見が、XサイエンスとCSや統計数理の両方から独立した新研究分野として第3世代Xインフォマティクスへと展開し、更にはイン・シリコ・サイエンスを目指す第4世代へと成熟する。マテリアルズ・インフォマティクスの第1世代は2011年の米国でのMGI (Materials Genome Initiative) の立ち上げと、2015年の日本での情報統合型物質・材料開発イニシアティブ (MI<sup>2</sup>) の発足に象徴される。DFT計算で得られた特定カテゴリの材料に関する均質データ集合に対して回帰推定を適用し、未知材料のDFT計算を回帰推定で置き換え、パラメータ値の最適化を図る試みや、少ない実験データ集合から始めて、ベイズ最適化を用いて実験データを徐々に効率よく増やして最適材料を求める試みなどにおいて、初期の成功を示してきた。一方で、機能無機結晶材料探索においては、対象データ集合が不均質で、回帰分析の前に適切に分割を行う必要があることが認識されてきた。分割のためのクラスタリングやパターン・マイニングに使えるような記述子をどのように系統的に定義すべきか未だによく解っていない。分割後の均質データ集合は小さく、過学習を避けるには回帰推定に精々5~7個程度の説明変数しか用いることができない。無機結晶材料の適切な記述子の定義や、これを用いたデータ集合の適切な分割や回帰推定は今後の重要な課題となっている。また、革新的人工機能材料を目指し、機械学習に適したデザイン可能性を持つナノ構造アーキテクチャの研究も今後の重要課題である。