

材料開発における4つのインフォマティクス

岩崎 悠真 (国研)物質・材料研究機構 マテリアル基盤研究センター

1. はじめに

磁性材料や触媒材料など、自動車に使われる材料は年々複雑になっており、近年の材料開発の難易度は非常に高い。この難化した材料開発を効率的に進めるべく、人工知能や機械学習などいわゆるデータ科学の活用が進められており、新規材料の予測から材料合成プロセスの最適化、計測効率化、そして材料メカニズムの理解に至るまで、様々な分野でデータ科学が重要な役割を果たしている^{1,2)}。この分野は、新規物質をデータ科学で予測する『マテリアルズ・インフォマティクス』、材料合成プロセスの最適化・可視化を目的とする『プロセス・インフォマティクス』、材料計測・解析を効率化する『計測インフォマティクス』、材料メカニズムの理解を深める『物理インフォマティクス』に分けて考えることができる(※ただし、上記4つの技術を全部含めて『マテリアルズ・インフォマティクス』と表現することもある)。

これら4技術は独立ではなく、図1のようにそれぞれがつながってループを組むような関係にある。『予測・発見する(マテリアルズ・インフォマティクス)』⇒『合成する(プロセス・インフォマティクス)』⇒『計測・解析する(計測インフォマティクス)』⇒『理解する(物理インフォマティクス)』⇒『予測・発見する(マテリアルズ・インフォマティクス)』⇒・・・のループである。この『予測』⇒『合成』⇒『計測』⇒『理解』⇒『予測』⇒・・・の流れは、多くの材料開発に当てはまる。

本記事では、上記4つの技術を簡単に説明する。

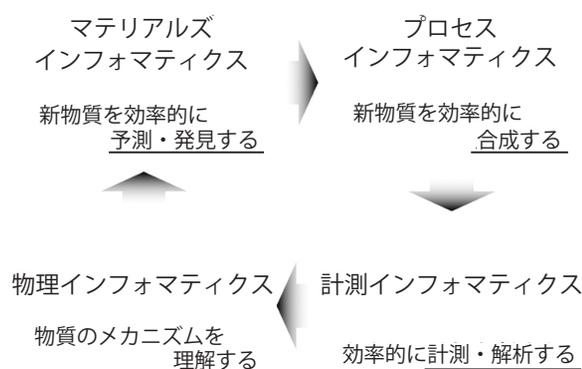


図1 材料開発における4つのインフォマティクス

2. マテリアルズ・インフォマティクス

人工知能や機械学習の技術を用いて新規物質を予測する『マテリアルズ・インフォマティクス』の研究は、2010年前半頃から活発に行われてきた。材料の物性値、結晶構造、組成などのデータから、様々な新規物質の組成・構造の予測が可能となっている。すでに書籍やレビュー論文も多数出版されている^{3,4)}。

例えば近年、Google DeepMindの研究者らのマテリアルズ・インフォマティクス関連の研究がNature誌から発表され話題となった⁵⁾。マテリアルズ・インフォマティクスの登場により、従来の材料研究者以外にも、本格的に材料開発の分野に参入することが可能となったのである。実際にGoogle以外にも、従来では材料開発に直接関係してこなかった企業が続々とマテリアルズ・インフォマティクスの分野に参入してきており、材料開発の在り方が変わりつつある。

3. プロセス・インフォマティクス

前述したマテリアルズ・インフォマティクスの技術発展により、新規材料の組成・構造の予測はデータ科学からある程度可能となってきた。しかし、その予測された材料を実際に合成することは簡単ではない。材料合成プロセスによって、出来上がる物資・材料は大きく変化するからである。材料特性が良いであろう組成・構造をマテリアルズ・インフォマティクスで見つけたが、どう頑張ってもその材料を合成することはできなかった、ということは頻繁に起こる。

材料を合成するプロセスのパラメータは、温度・圧力・時間・使用する装置・・・など、非常に多い。これら材料合成プロセスをデータ主導で最適化することがプロセス・インフォマティクスの主な目的である。

マテリアルズ・インフォマティクスに比べると、プロセス・インフォマティクスは難しいと考えられている。その主な理由はプロセスに関するデータがあまり蓄積されていないことである。マテリアルズ・インフォマティクスで使うことができる材料の物性に関わるデータ蓄積は、世界中で進められている。例えば NIMS の MatNavi という材料データベース群には様々な材料に関するデータが蓄積されており、日々どんどん増えている。一方、材料の作り方（プロセス）に関する有用なデータベースは、現在あまり多くない。特にインフォマティクスの観点からは失敗データ、つまり目的の材料が合成できなかった時のプロセス条件のデータが非常に重要となるが、これら失敗データをしっかりと記録して蓄積してあるデータベースは非常に少ない。そのため、現在のプロセス・インフォマティクスは、『プロセスデータを蓄積すること』および『少ないデータで材料開発プロセスを最適化（自律化・可視化）すること』が重要となる⁶⁻¹¹⁾。

4. 計測・インフォマティクス

材料合成を終えたら、次はその材料の物性測定・評価を行う。この物性測定・評価をデータ科学の力を借りて効率的に行うのが計測インフォマティクスである。例えば、2017年にクライオ電子顕微鏡法 (Cryo-EM)^{12,13)}の

開発者たちがノーベル化学賞を受賞した。この手法は電子顕微鏡技術と統計的データ処理技術が融合している手法であり、一種の計測インフォマティクスといえる。特に近年、ハイスループット実験が盛んにおこなわれるようになり、人間では処理できないほどの大量の計測データが瞬時に手に入るようになった。例えば放射光施設では、二次元スペクトルマッピングなどのデータが手軽に取得できるようになり、機械学習の出番が非常に増えている。限られた時間の中で（例えば放射光施設の利用時間内に）効率的に測定するために、測定するサンプルや位置や測定パラメータをデータ主導で決定したり（自律的計測）¹⁴⁻¹⁶⁾、得られた大量のデータを高速に解析したりすること¹⁷⁻²³⁾が、計測インフォの主な仕事となる。

計測インフォマティクスが求められている背景には、上記のような『効率的な計測・解析』だけでなく、『解析結果の客観性』もある。例えば、スペクトルデータをフィッティングするときのことを考えてみる。スペクトルデータのフィッティングはある種の職人技と言われており突き詰めると非常に奥が深く難しい。そのため、フィッティングを行う人によって結果が違うということが多々起こる。人間の事前知識を用いず、データ主導でフィッティングを行った場合の解析結果は、解析する人に依存しないので、客観性の高い解析結果であると言える。

5. 物理インフォマティクス

単純に新規材料・物性を“予測”するのではなく、その材料物性・材料開発プロセス等のメカニズムを“理解”するためにデータ科学を活用する技術領域は物理インフォマティクスと呼ばれる（※ただし、物理インフォマティクスの守備範囲は非常に広く、材料物性だけでなく、宇宙物理、量子コンピューティング、超紐理論等への機械学習の応用も物理インフォマティクスと言われる）。物理インフォマティクスの領域では、深層学習のようなブラックボックス型の機械学習だけでなく、モデル解釈性の高いホワイトボックス型の機械学習も頻繁に使われる。解釈性の高い機械学習モデル (Interpretable machine learning や Explainable AI など) と呼ばれてい

る²⁴⁻²⁷⁾を用いることによって、機械学習モデルの内部を人間が物理・化学・材料学の知見をベースに考察できるようになるため、材料物性やプロセスの“理解”を深めることが可能となる。実際に様々な分野の材料開発ですでに使われている²⁸⁻³¹⁾。材料物性やプロセスに関する理解を深めることができれば、マテリアルズ・インフォマティクスやプロセス・インフォマティクスへのより効率的なフィードバックを行うことが可能となる。例えば、材料物性に関する知識が深まれば材料探索空間を再定義したり限定したりすることができ、プロセスに関する知識が深まればプロセス空間を限定してより高速な材料創製が可能となる。

6. おわりに

物質・材料研究の分野における4つのインフォマティクス技術について簡単に説明した³²⁾。データ科学を駆使した材料開発分野を読者が俯瞰するうえで分かりやすい分類となっていれば幸いである。ただし、それぞれの定義や境界線は結構曖昧になっている(日々変化している)ため、その点ご注意ください。

文献

- 1) 岩崎 悠真, “マテリアルズ・インフォマティクス 材料開発のための機械学習超入門” 出版社: 日刊工業新聞社 (2019)
- 2) 岩崎 悠真, “マテリアルズ・インフォマティクス II 機械学習を活用したマテリアル DX 超入門” 出版社: 日刊工業新聞社 (2022)
- 3) R. Batra et al. Emerging materials intelligence ecosystems propelled by machine learning. *Nat. Rev. Mater.* 1-24 (2020)
- 4) R. Ramprasad et al. Machine learning in materials informatics: recent applications and prospects. *npj Comput. Mater.* 3, 54 (2017)
- 5) A. Merchant et al. Scaling deep learning for materials discovery. *Nature* 624, 80-85 (2023)
- 6) R. Shimizu et al. Autonomous materials synthesis by machine learning and robotics. *APL Materials* 8, 111110 (2020)
- 7) B. Burger et al. A mobile robotic chemist. *Nature* 583, 237-241 (2020)
- 8) N. J. Szymanski et al. An autonomous laboratory for the accelerated synthesis of novel materials. *Nature* 624, 86-91 (2023)
- 9) P. Nikolaev et al. Autonomy in materials research: a case study in carbon nanotube growth. *npj Comput. Mater.* 2, 16031 (2016)
- 10) C. W. Coley et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* 365, eaax1566 (2019)
- 11) Z. Li et al. Robot-accelerated perovskite investigation and discovery. *Chem. Mater.* 32, 5650 (2020)
- 12) R. Fernandez-Leiro et al. Unravelling biological macromolecules with cryo-electron microscopy. *Nature* 537, 339-346 (2016)
- 13) E. Nogales. The development of cryo-EM into a mainstream structural biology technique. *Nat. Methods* 13, 24-27 (2016)
- 14) M. N. Marcus et al. A Kriging-Based Approach to Autonomous Experimentation with Applications to X-Ray Scattering. *Sci. Rep.* 9, 11809 (2019)
- 15) A. G. Kusne et al. On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat. Comm.* 11, 5966 (2020)
- 16) T. Ueno et al. Automated stopping criterion for spectral measurements with active learning/ *npj Comput. Mater.* 7, 139 (2021)
- 17) A. Arima et al. Selective detection of single-viruses using solid-state nanopores. *Sci. Rep.* 8, 16305 (2018)
- 18) G. Imamura et al. Development of machine learning models for gas identification based on transfer functions. *Proc. 17th International Meeting on Chemical Sensors, Vienna* (2018)
- 19) T. Matsumura et al. Spectrum adapted expectation-maximization algorithm for high-throughput peak shift analysis. *Sci. Technol. Adv. Mater.* 20:1, 733-745 (2019)
- 20) H. Shinotsuka et al. Development of spectral decomposition based on Bayesian information criterion with estimation of confidence interval.
- 21) M. Shiga et al. Sparse modeling of EELS and EDX spectral imaging data by nonnegative matrix factorization. *Ultramicroscopy* 170, 43-59 (2016)
- 22) A. Baliyan et al. Machine Learning based Analytical Framework for Automatic Hyperspectral Raman Analysis of Lithium-ion Battery Electrodes. *Sci. Rep.* 9, 18241 (2019)
- 23) S. Masubuchi et al. Deep-learning-based image segmentation integrated with optical microscopy for automatically searching for two-dimensional materials. *npj 2D Mater. Appl.* 4, 3 (2020)
- 24) C. Rudin et al. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* 1, 206-215 (2019)

- 25) P. Voosen. The AI detectives. *Science*, 357, 6346 22-27 (2017)
- 26) A. Adadi et al. Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6, 52138-52160 (2018)
- 27) R. Guidotti et al. A Survey of Methods for Explaining Black Box Models. *ACM Comput. Serv.* 51, 5 (2018)
- 28) Y. Iwasaki et al. Identification of advanced spin-driven thermoelectric materials via interpretable machine learning. *npj Comput. Mater.* 5, 103 (2019)
- 29) Y. Suzuki et al. Symmetry prediction and knowledge discovery from X-ray diffraction patterns using an interpretable machine learning approach. *Sci. Rep.* 10, 21790 (2020)
- 30) J. Jiménez-Luna et al. Drug discovery with explainable artificial intelligence. *Nat. Mach. Intell.* 2, 573-584 (2020)
- 31) J. Feng et al. Explainable and trustworthy artificial intelligence for correctable modeling in chemical sciences. *Sci. Adv.* 6, 42, eabc3204 (2020)
- 32) 岩崎 悠真, データサイエンスと材料開発, プロセスインフォマティクス, 出版社: 情報機構 in press