

ここまで来たマテリアルズインフォマティクスの最前線

Introduction to Materials Informatics and MI²I Project

計算科学振興財団¹、兵庫県立大学² 伊藤 聡^{1,2}

Foundation for Computational Science¹, University of Hyogo² Satoshi Itoh^{1,2}

E-mail: itoh@j-focus.or.jp

データ駆動型科学の一分野としてマテリアルズインフォマティクス (MI) が著しい進歩を見せているが、物質科学の分野におけるデータの在り方は他の分野と異なるところがある。ハンドブックやデータ集に収録されている物性データは信頼性の高いものであるが、機械学習を適用するにはデータ量が少なすぎるのである。2015 年から 5 年間、実施した JST 事業「情報統合型物質・材料開発イニシアティブ (MI²I)」では、この課題に対していくつかの試みを行い、一定の成果が得られたので、これを紹介したい。

第一のアプローチは大量にデータを創出できるシミュレーション技術の活用である。分子構造・結晶構造を与えれば、その電子構造を計算することは容易であり、その結果から金属／非金属の判定や光吸収スペクトル、熱伝導特性や電子伝導特性の予測などが可能である。このようにして作られたデータセットは、不純物など影響は考慮していないとはいえ、機械学習による材料開発には大いに役立つので、現在、世界中でこうしたデータベースの構築が行われている。また、計算データを補完する、あるいは検証する形で実験データの収集も進んでおり、これらデータを FAIR 化して機関間にわたる活用を想定した大規模レポジトリシステムも開発されている。

第二は機械学習手法そのものの進展が挙げられる。スモールデータ問題は機械学習にとっても新たな課題であった。この問題に対して、別のデータセットを用いた事前学習結果を利用する転移学習手法が提案されている。実際、非常に少ないデータ数から予測モデルを構築できる事例が報告されている。また、同じ物性に関するデータセットであれば、融合して使えばデータ数が増えるが、異なるデータセットでは収集条件が異なるので (たとえば測定温度など)、単純に融合させると非常に質の悪いデータセットなる。これに対しては数値そのものではなく、データ間の関係をもとにしたサロゲートモデルを経由する機械学習手法 (プリファレンス学習) が提案されている。

以上のような取り組みによって MI²I では高熱伝導性高分子材料、圧力誘起超伝導材料、波長選択熱輻射多層膜などを探索し、かつ実際に合成してその特性を確認した。これは手法の正しさや有効性を実証するものであり、機械学習はとくに組成最適化や元素選択の組み合わせ探索に関しては大きな成果を収めつつあるが、現時点では材料研究者が思いもよらなかった優れた特性を示す新材料、新規組成材料が開発された事例はまだない。しかながら、この分野は急速に発展している技術分野であり、マテリアルズインフォマティクスをはじめとするデータ駆動型のアプローチが材料開発の現場を大きく変えていくことを期待したい。