

機械学習による化学反応の予測と設計

理化学研究所 革新知能統合研究センター/北海道大学 化学反応創成研究拠点 瀧川 一学

1. はじめに

化学反応は物質を別の物質へ変換するための基礎として私たちの生活を支えている。エネルギー・電子材料・医薬・化学産業・農業など幅広い分野の屋台骨であるだけでなく、生命現象や私たち自身の存在も化学反応により成立している。こうした重要性にも関わらず、化学反応の予測・設計・探索・制御は現在も経験的な試行錯誤とセレンディピティに依存しており、属人的・偶発的な方法論からデータに基づく実験計画による合理的・効率的な方法論への転換が模索されている。

2. 機械学習の課題

このような強いニーズは従来の機械学習の限界を明らかにするとともに、多数の興味深い技術課題を生み、転換点にあると言われる人工知能分野の文脈でも新たな潮流を形成しつつある(瀧川, 2019)。本発表ではその中から主に下記の3点に関する課題と現状を簡単に概観する。

課題 1. Exploration と Out-of-distribution 予測. 機械学習モデルは与えられた訓練データの代表であるが、化学では今までに(データに)無いものの探求が関心である。従って、訓練データ分布の外の予測(OOD 予測)やデータをさらに取得する適応的実験計画をどのように実現するかが重要である。入力空間では外挿的な問題を潜在空間で内挿的にするための表現学習は一つの鍵である。また、「新しいことの学習」に伴うデータの活用と探索のトレードオフは機械学習の基本であるが、分子構造の逐次的な部分置換やロボットによる自動合成系などの逐次的意思決定に現れる強化学習やモデルベース最適化を念頭に、特に「探索(exploration)」の問題が研究されている。

課題 2. Compositionality・組合せと Transferability. データが多量にあり運用時に起こり得る事例に近いものが満遍なくカバーされていれば従来の機械学習で十分であるが、化学では現実的な制約からこれを期待できない。現行の機械学習は自然科学用途ではあまりに data hungry であり sample efficient な手法が必要不可欠である。人間は「言語を操り、虚構を想像し、認識する」稀有な能力を持ち、このことが未経験(zero-shot)や数回の経験(few-shot)の状況でも有効な学習を実現している。例えば、対象を何らかの汎用性のある要素の組合せと捉え新たな組み合わせを自在に構成する compositionality, 複数の少数例から何を学習すべきかを獲得する meta learning、学習時に明示的な教師ラベルを与えられずに行う self-supervised learning やその transferability, 要素数の異なる集合・系列・グラフ等の組合せ的な離散構造の機械学習、等が研究されている。

課題 3. 演繹と帰納の融合. 化学の問題は電子の状態に関する方程式(Schrödinger 方程式)で説明でき、近年の計算基盤の進展により化学反応も計算化学・シミュレーションの対象となっている。しかし、多体問題により陽な求解はできず近似の仮定の例外があり得る、計算できる系の規模が限られる、高忠実度計算には非実用的な時間がかかる、計算で考慮されない外乱要因が多数ある、現実の反応はモル単位の多数の分子で起こる、など第一原理が分かれば反応が自由自在に制御できるわけではない。従って、第一原理的な演繹と機械学習的な帰納を融合する様々なアプローチが研究されている。人工知能分野でも「データにないこと」の検証や planning(計画)に重要となる演繹型の記号的な推論・計画立案と帰納型の機械学習の融合は主要な技術課題の一つである。

参考文献

瀧川 一学 (2019) 人工知能基本問題研究会 (FPAI). 人工知能, **34**(5), 603–611.
<http://id.nii.ac.jp/1004/00010296/> (Open Access)