

転移学習によるMI技術の発展に期待

◆ 転移学習を可能にするデータベースが公開

2019年10月、統計数理研究所と物質・材料研究機構の共同研究グループはマテリアルズインフォマティクス（MI）のオープンソースプラットフォームXenonPyのライブラリを公開した。ライブラリには開発した低分子、高分子、無機材料の45種類の特性を対象にした約140,000個の機械学習の予測モデルが含まれている。このモデルを転移学習という手法を用いて、個別の企業や研究機関が所有するデータと組み合わせることで、材料設計や物性予測に活用することができる。

転移学習は、少ないデータから高精度な予測をするために、似た特性の予測モデルを転用する技術である。画像認識や言語処理の領域では広く活用されており、例えば画像から犬の種類を予測するモデルでは、犬の画像を大量に集めなくとも別の学習済み画像予測モデルを転移学習することで、精度の高い予測ができる。

研究グループはライブラリを利用した熱伝導率予測において、19の高分子熱伝導率データに対し、高分子のガラス転移点や低分子化合物の比熱容量などの訓練済み予測モデルを転移させ、データには含まれていないメソゲン基を骨格とする高分子の物性を予測した。同様に、無機化合物では熱伝導率が370W/mK以内のデータ群から、15～3371W/mKの範囲で単結晶の熱伝導率を予測し、転移学習の有効性と、MIでは困難とされていた外挿的な物性予測が可能であることを示した。

◆ 既存のデータベースを転移学習に活用

19年11月、ノースウェスタン大学の研究グループは、密度汎関数理論（DFT）計算による材料予測において、既存の大規模DFT計算データベースを転移学習させ、実験データと組み合わせる手法を発表した。この転移学習と実験データを統合した予測モデルによって、計算データのみを機械学習させたモデルよりも予測誤差を抑え、より高精度な材料の生成エネルギーの推定を可能にした。

MIの課題として、データを取得する実験や計算に多くの時間やコストがかかることや、外挿領域の予測が難しいことが指摘されているが、転移学習を用いた取り組みにより、これらの課題を克服することが期待されている。 【塚原祐介】