

小規模な実験データベースの機械学習に基づく有機分子の物性の外挿予測モデルの構築

(早大理工) ○ 畠山 歓・小柳津 研一

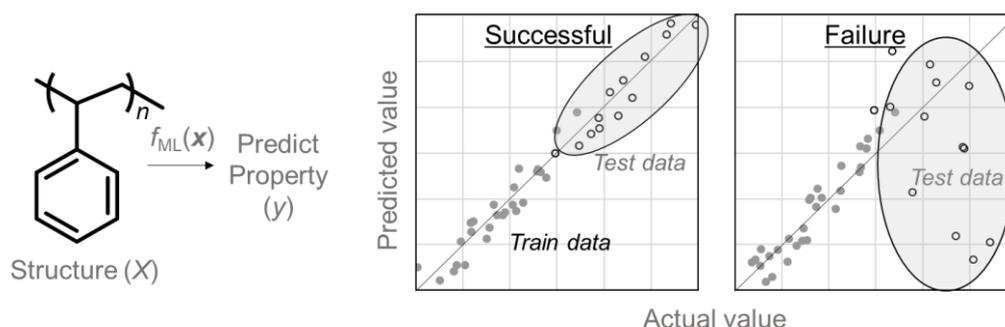
Extrapolating molecular properties using machine learning models trained with small-size experimental databases (*Department of Applied Chemistry, Waseda University*) ○ Kan Hatakeyama-Sato, Kenichi Oyaizu

Materials informatics enables the estimation of material properties from their structures using machine learning models. Still, the prediction has been difficult, especially when small databases with missing data are used and when so-called extrapolation is pursued. Here, we introduce imputation techniques to cope with the issues. The demands of descriptors and models for the prediction will also be discussed in the presentation.

Keywords : *Materials Informatics, Organic Functional Materials, Machine Learning*

マテリアル・インフォマティクスの方法論の発展により、機械学習関数 f_{ML} をもとに材料構造(x)から所望の物性 y を予測する手法が開発されつつある^{1,2}。一方で未解決の課題として、主に実測データベースを扱う際の 1)学習データの件数不足、2)欠損値への対処や、予測したい物性が訓練データの範囲外にある 3)外挿予測時の精度低下などが残っている。

本報告では代入法(imputation)を回帰問題に応用し、小規模な実験データベース条件下での予測精度の改善を試みた。密度や分解温度等の物性値を予測するタスクにおいて、異なる物性データベースを統合し、更に欠損値を機械学習モデルで代入しながら回帰することで、多くの物性(>70%)で予測性能を上げられることが分かった。当日は内挿・外挿予測の為の学習モデルや記述子の選定法等についても、併せて議論する。



1 R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, *Npj Comput. Mater.* **2017**, *3*, 54.

2 K. Hatakeyama-Sato, K. Oyaizu, *Commun. Mater.* **2020**, *1*, article number: 49.