

機械学習による HPLC 分析におけるフェノール類の保持時間の予測

(同志社大生命) ○中村 祐士・余 光夫・大江 洋平・太田 哲男

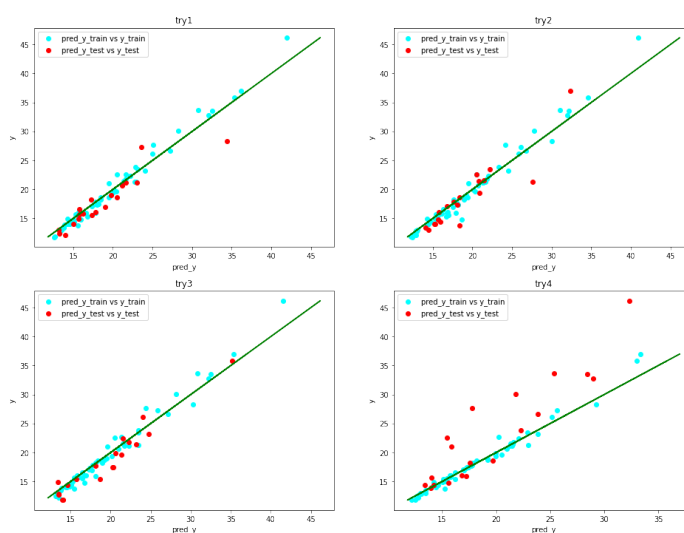
Machine Learning-Based Retention Time Prediction of Phenols in HPLC Analysis (¹*Faculty of Life and Medical Sciences, Doshisha University*) ○Yushi Nakamura, Koufu Sha, Yohei Oe, Tetsuo Ohta

We investigated retention time (RT) prediction for phenols in high pressure liquid chromatography (HPLC) analysis. RT prediction model of phenols was built using random forest as a machine learning algorithm and its prediction performance was estimated by 4-fold cross validation (R-squared average score: 0.79, MAE average score: 2.07).

Keywords : Chemoinformatics, HPLC, Machine Learning, Phenols

HPLC 分析は極めて優れた分離分析の一つであり、有機合成研究における合成サンプル分析手法の一つとしてだけでなく、品質管理や環境測定等といった幅広い分野において利用されている。HPLC 分析における各サンプルの保持時間を予測することができれば、分析条件やカラムの決定に有用な知見と成り得るため、保持時間予測モデルの構築に関する研究が古くから行われている。¹⁾ 本研究では、フェノール類をターゲットとした機械学習によるそれらの保持時間の予測について検討した。

具体的には、研究室で保有する 76 種類のフェノール類について、カラムとして Wakopak® Wakosil®-II 5C18HG (4.6×100 mm) を用いた際の以下の分析条件における保持時間の予測について検討した (Flow rate: 0.1 mL/min, Solvent: CH₃CN/H₂O = 60/40, Temp.: 40°C). RDKit, Mordred, alvaDesc から生成した化学記述子について、相関関係を基準に約 1000 個まで絞りこみ、それらを説明変数とした。目的変数を保持時間とし、上述の説明変数を用い、ランダムフォレストにより予測モデルを構築し、k 分割交差検証 (k = 4) によりその予測精度を評価した (右図)。k 分割交差検証における R₂ 値の平均値は 0.79, MAE の平均値は 2.07 となり比較的良好な予測モデルが構築できることが分かった。また、重要度の高い変数として、logP のような保持時間への影響が良く理解されている変数だけでなく、分子グラフに由来する変数も複数確認された。詳細については本会で報告する。



1) P. R. Haddad, M. Taraji, R. Szucs, *Anal. Chem.* **2021**, 93, 228.