

多環芳香族の接触分解反応における機械学習を用いた生成物組成予測

(信州大) ○嶋田 五百里・長田 光正・福長 博・古山 通久

1. 緒言

近年、機械学習を用いた固体触媒反応の生成物予測や触媒設計が活発に行われている¹⁾。これまでの研究は C1 化学等の比較的単純な反応系への適用例が多いが、重質油精製などの原料組成や反応機構が複雑な反応系への適用も期待される。本研究では、多環芳香族炭化水素の接触分解反応を対象とし、機械学習を用いた生成物組成予測に取り組んだ。この反応系では、多環芳香族は単独では非常に安定であるが、共存物質（パラフィンやナフテン）との相互作用により分解が進行することが示唆されている^{2,3)}。一方で、複雑な反応系の中から重要な相互作用を抽出することは困難であるため、その解決方法として機械学習の活用を検討した。

機械学習を用いた予測の課題として、予測精度と解釈可能性のトレードオフの関係が挙げられる。反応機構の解明や重要な因子の抽出に機械学習モデルを活用するためにはモデルの解釈可能性が高いことが求められるが、一般に解釈可能性の高いモデルは予測精度に劣る。そこで本研究では、解釈可能性の高い線形回帰を用いつつも、物理化学に基づく特徴量エンジニアリングを行うことで予測精度の向上を試みた。

2. 解析

反応温度、接触時間、原料 (9-メチルアントラセン (3 環芳香族)、*n*-ヘキサデカン (パラフィン)、テトラリン (ナフテン)、トルエン (不活性物質)、9,10-ジヒドロアントラセン (反応中間体)) の混合比を変更して行った 62 点の接触分解反応データ^{2,3)}を解析対象とした。反応条件及び原料組成から特徴量を作成し、単環芳香族生成量を目的変数として、線形回帰 (LASSO) を用いた予測モデルを作成した。モデルの検証には 10-fold クロスバリデーションを用い、テストデータの二乗平均平方根誤差 (RMSE) 及び決定係数 (R^2) を比較した。

3. 結果および考察

本研究で考慮した特徴量を Table 1 に示す。基本特徴量 (反応温度、接触時間、各原料の入口濃度) に加え、原料間の相互作用を表現する濃度交差項を組み込んだ。さらに、物理化学に基づく特徴量エンジニアリングとして、温度をアレニウス型 ($\exp(-E_a/RT)$) で表現する、反応速度項や微分反応器を仮定した濃度変化項を導入する、などを行い、物理化学的な情報を取り入れた特

Table 1 Descriptors used in model construction

Basic descriptors	
Temperature	T
Contact time	t
Feedstock concentration	$X_i (i = 1 - 5)$
Additional descriptors (Physics-based feature engineering)	
Concentration interaction	$X_i X_j (i, j = 1 - 5)$
Arrhenius type temperature	$\exp(-E_a/RT)$
Reaction rate	$\exp(-E_a/RT) \cdot X_i$
Concentration change	$\exp(-E_a/RT) \cdot X_i X_j$
	$\exp(-E_a/RT) \cdot X_i t$
Concentration change	$\exp(-E_a/RT) \cdot X_i X_j t$

Table 2 RMSE and R^2 (mean $\pm$ SD) with different models

Model	RMSE	R^2
LASSO	0.796 $\pm$ 0.316	0.849 $\pm$ 0.497
Support vector regression	0.477 $\pm$ 0.283	0.956 $\pm$ 0.064
Random forest	0.581 $\pm$ 0.260	0.930 $\pm$ 0.106
Exhaustive model	0.448 $\pm$ 0.260	0.955 $\pm$ 0.183
Physics-based model	0.421 $\pm$ 0.217	0.966 $\pm$ 0.048

徴量を作成した。これを用いて、LASSO による線形回帰モデル (Physics-based model) を作成した。また、物理化学に基づく特徴量エンジニアリングとの比較のため、基本特徴量のすべての組み合わせにおける 2 次及び 3 次の項を網羅的に足し合わせた LASSO 回帰モデル (Exhaustive model) も作成した。

作成したモデルの予測精度を、基本特徴量のみを用いた線形回帰 (LASSO) 及び非線形回帰 (サポートベクター回帰及びランダムフォレスト) と比較した。それぞれのモデルにおける RMSE 及び R^2 を Table 2 に示す。非線形回帰手法や Exhaustive model よりも Physics-based model の方が高い予測精度を達成した。さらに、Physics-based model では、影響の大きな特徴量として 3 環芳香族とナフテンとの 2 次反応速度や反応中間体の分解による濃度変化を表す項など、解釈可能な要素が選択される傾向があった。これらの結果より、物理化学に基づく特徴量エンジニアリングによって解釈可能性と予測精度の高いモデル構築の可能性が示された。

謝辞

本研究は JSPS 科研費 20K05207 及び (一財) 石油エネルギー技術センターの革新的石油精製技術のシーズ発掘事業の支援を受けた。

- 1) T. Toyao, Z. Maeno, S. Takakusagi, T. Kamachi, I. Takigawa, K. Shimizu, *ACS Catal.*, **10**, 2260 (2020).
- 2) I. Shimada, K. Takizawa, H. Fukunaga, N. Takahashi, T. Takatsuka, *Fuel*, **161**, 207 (2015).
- 3) 嶋田, ペトロテック, **40**, 81 (2017).