

2024年度 JPECフォーラム

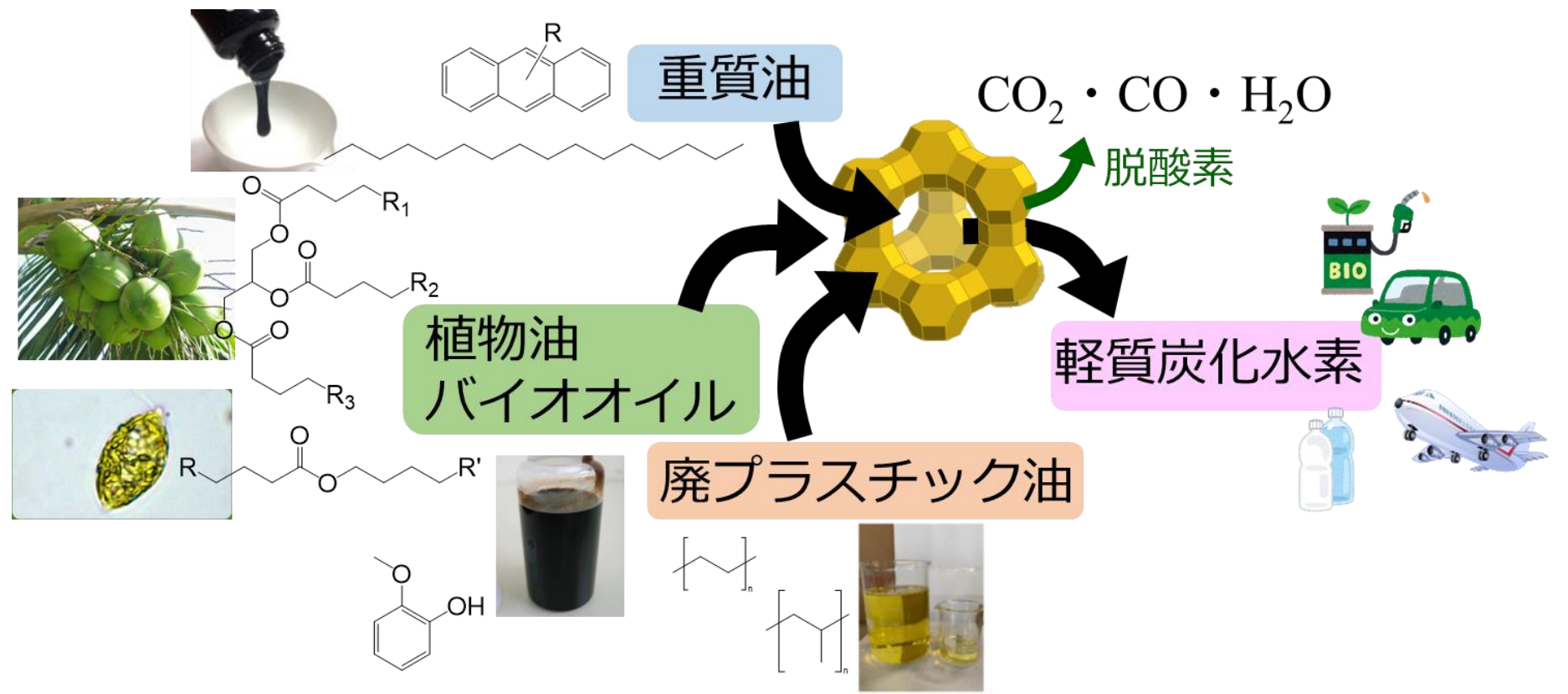
バイオオイルと重質油の混合接触分解における
生成物組成を予測する機械学習モデルの構築

2024年5月14日

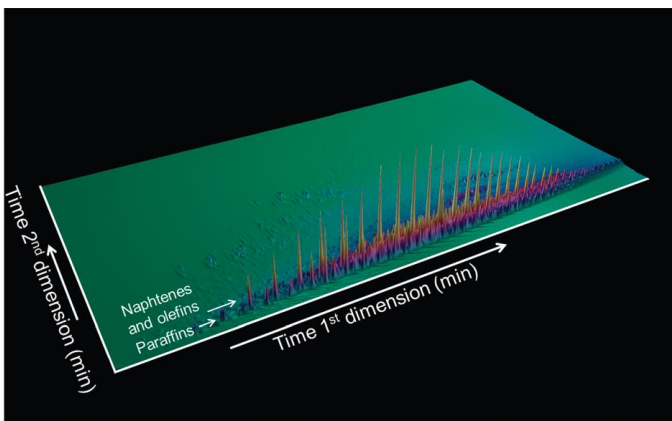
信州大学 嶋田五百里

流動接触分解 (FCC) プロセス

- 石油精製において、重質油を分解してガソリン・ナフサを生成するプロセス
- 従来の役割である重質油の分解に加え、再生可能資源やリサイクル資源の共処理 (co-processing) での利用も期待されている

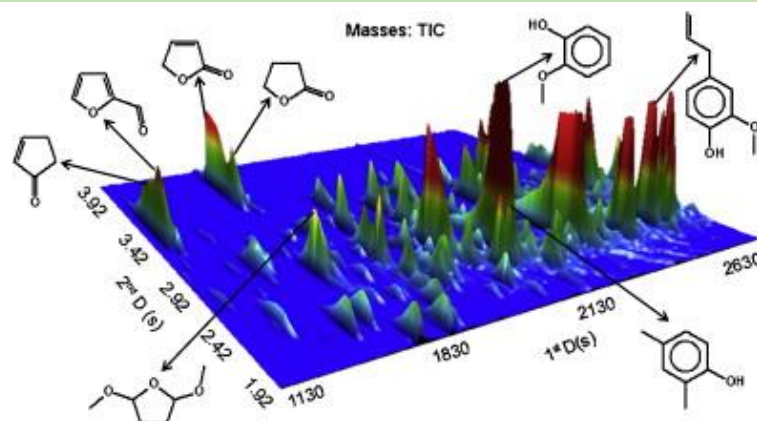


重質油



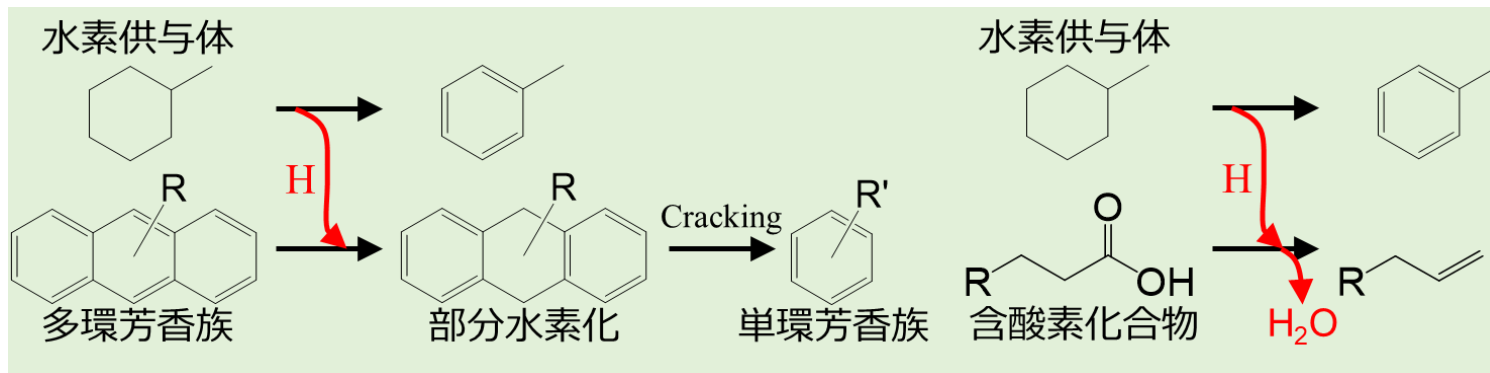
Vogt and Weckhuysen, *Chem. Soc. Rev.*, **44**, 7342-7370 (2015).

液化バイオマス (バイオオイル)



Tessarolo *et al.*, *J. Chromatogr. A*, **1279**, 68-75 (2013).

成分間での相互作用（特に水素の授受反応）が生成物に強く影響



Shimada *et al.*, *Fuel Process. Technol.*, **232**, 107267 (2022).

Shimada *et al.*, *J. Jpn. Petrol. Inst.*, **63**(1), 10-19 (2020).

極めて複雑な反応システムの中で

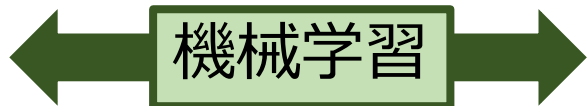
- どんな生成物がどのくらいできるか？
- 重要な反応因子は何か？



機械学習の利用

反応条件

- 反応温度
- 接触時間
- 原料組成

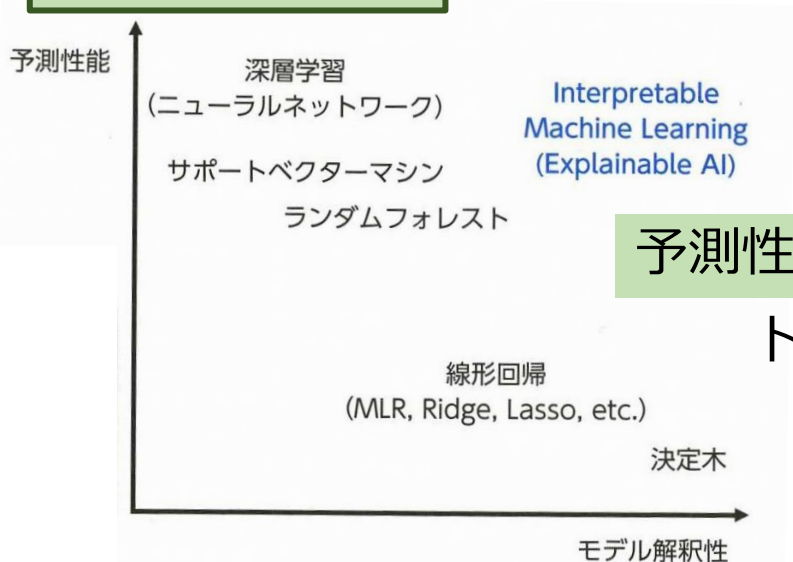


生成物組成

- 単環芳香族
- 重質成分
- H₂O/CO₂/CO など

生成物組成の予測
重要な反応因子の抽出

機械学習の課題



× ブラックボックスモデルでは重要な反応因子を抽出できない

予測性能 ⇔ 解釈可能性

トレードオフ

機械学習と物理の融合

Physics-informed machine learning

G. E. Karniadakis *et al.*, *Nat. Rev. Phys.*, **3**(6), 422–440 (2021).
S. Wang *et al.*, *Nat. Commun.*, **12**(1), 5288 (2021).

(岩崎悠真, マテリアルズ・インフォマティクス材料開発のための機械学習超入門)

本研究のアプローチ

物理化学 (反応速度論・反応工学) との融合により、
解釈可能性の高い機械学習モデル (線形回帰) の予測精度を向上させる

(一般的な) 特徴量エンジニアリング

基本特徴量

反応温度 T
 接触時間 τ
 原料濃度 X_i



T^2 $T \cdot \tau$ $T \cdot X_i$...
 T^3 $T^2 \cdot \tau$ $T \cdot X_i^2$...
 T/τ T/X_i ...
 \sqrt{T} $\sqrt{\tau}$ $\sqrt{X_i}$...

予測精度が上がる一方で
 解釈可能性は下がることも

物理化学に基づく特徴量エンジニアリング

Shimada et al., *Ind. Eng. Chem. Res.*, **62**, 21087-21099 (2023)

基本特徴量

反応温度 T
 接触時間 τ
 原料濃度 X_i

濃度交差項 $X_i \cdot X_j$

反応速度項

アレニウス型温度 $\exp(-E_a/RT)$
 1次反応 $\exp(-E_a/RT) \cdot X_i$
 2次反応 $\exp(-E_a/RT) \cdot X_i \cdot X_j$

E_a : ハイパーパラメータ

濃度変化項 (微分反応器モデル)

1次反応 $\exp(-E_a/RT) \cdot X_i \cdot \tau$
 2次反応 $\exp(-E_a/RT) \cdot X_i \cdot X_j \cdot \tau$

濃度変化項 (積分反応器モデル)

1次反応 $X_i \cdot \exp(-k\tau)$ ($k = A \exp(-E_a/RT)$)
 2次反応 $X_i^2 / (X_i + (k\tau)^{-1})$
 $X_i X_j (1 - \exp((X_j - X_i)k\tau)) / (X_i - X_j \exp((X_j - X_i)k\tau))$

逐次反応項

X_i / X_j
 $\exp(-E_a/RT) \cdot X_i / X_j$
 $\exp(-E_a/RT) \cdot X_i / X_j \cdot \tau$

回帰モデルの作成

LASSO

線形回帰モデルにおける縮小推定の一様

影響の小さいパラメータの係数をゼロにする傾向があり、特徴量選択が可能

$$\hat{y} = \sum_{m=0}^M w_m \phi_m = \mathbf{w}^T \boldsymbol{\phi} \quad \boldsymbol{\phi} = \begin{pmatrix} T \\ \tau \\ X_1 \\ \vdots \end{pmatrix}$$

$$\tilde{E}(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N_{\text{train}}} (y_{\text{train},n} - \mathbf{w}^T \boldsymbol{\phi}_{\text{train},n})^2 + \frac{\alpha}{2} \sum_{m=1}^M |w_m|$$

→最小化する \mathbf{w} を探索

\hat{y} : 目的生成物の生成量の予測値

\mathbf{w} : パラメーター

M : 特徴量の数

$\boldsymbol{\phi}$: 特徴量

N_{train} : 訓練データ数

α : ハイパーパラメーター

Physics-based FE (feature engineering) model

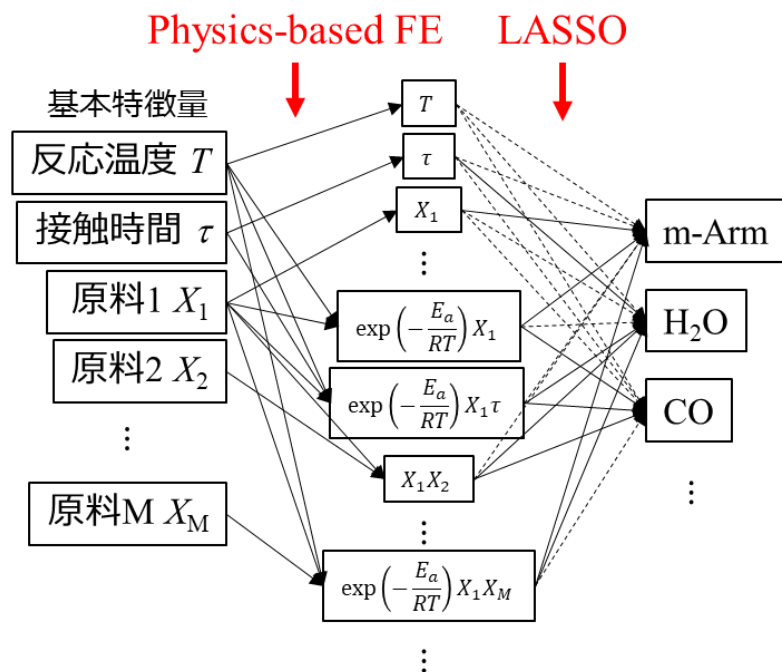
物理化学に基づく特徴量エンジニアリング

全ての成分間の反応を表す記述子を作成



LASSO回帰による特徴量選択

データに基づき重要な記述子を選択



構築したデータセット

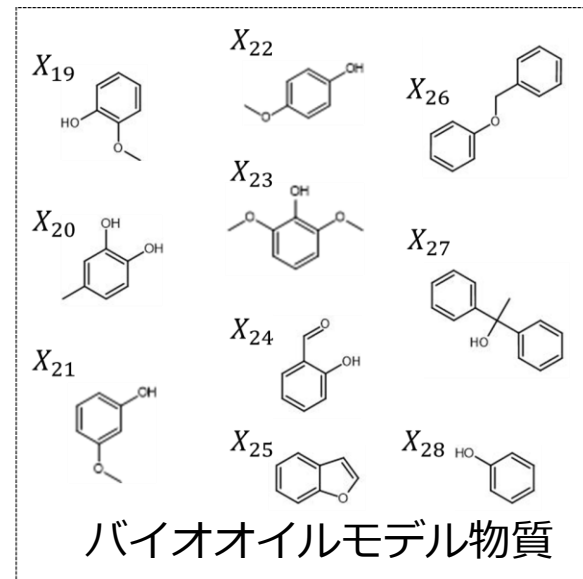
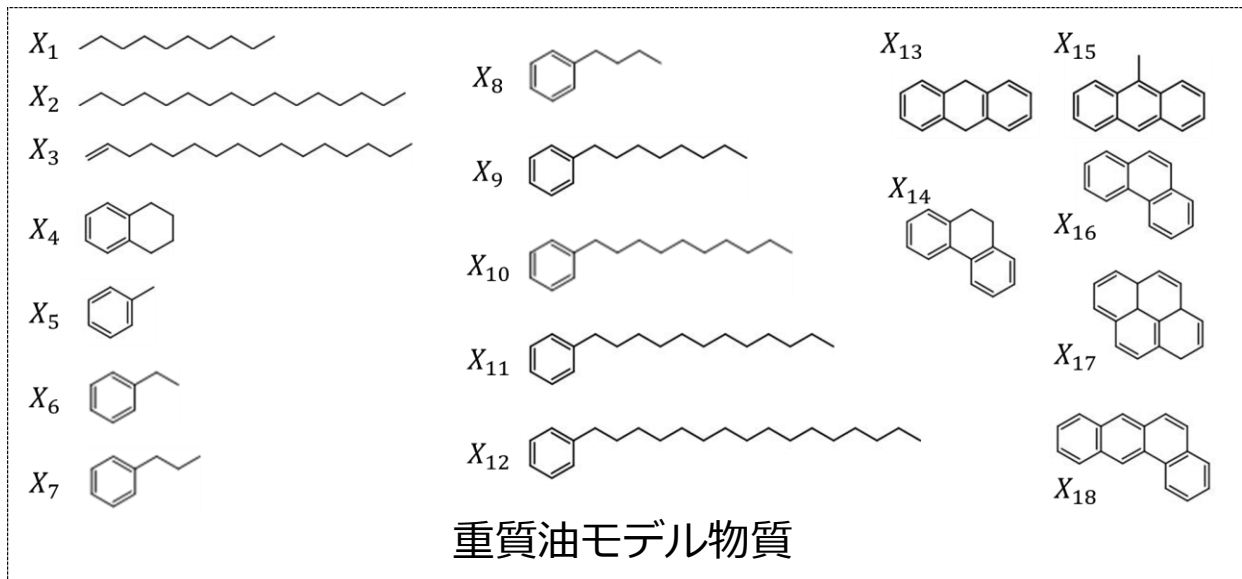
実験データ (257点)

基本の特徴量

反応温度 450~550°C

接触時間 0.16~7.50 g/g

原料組成 $X_i (i = 1 \sim 28)$



モデル構築

生成物中の各成分の含有量を予測

単環芳香族類 H_2O CO CO_2
フェノール類 ベンゼンジオール類

257

データセット



247

トレーニングデータ

10

テストデータ

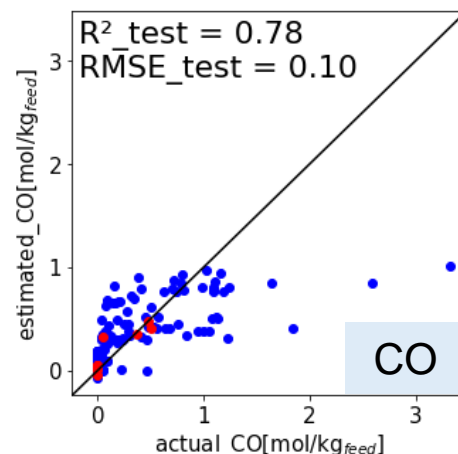
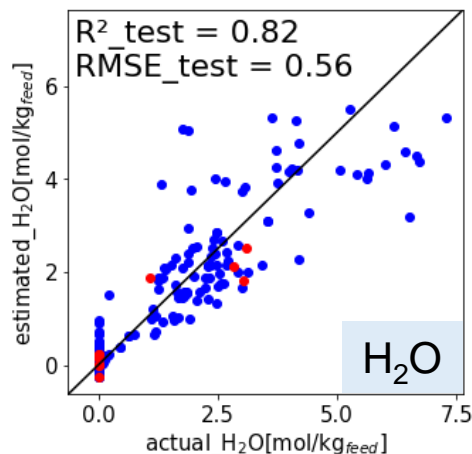
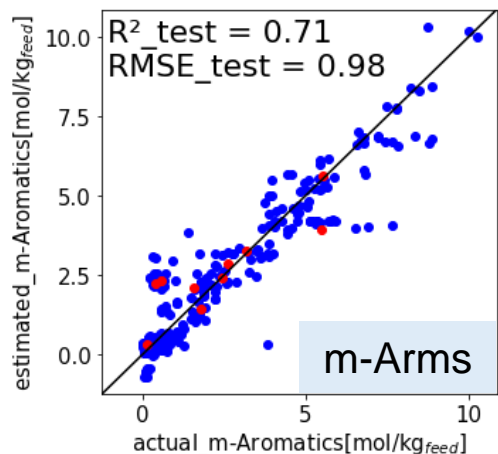
(モデル構築) (予測精度の評価)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

y_i ; 実測値 \bar{y} ; 平均値
 \hat{y}_i ; 予測値 n ; データ数

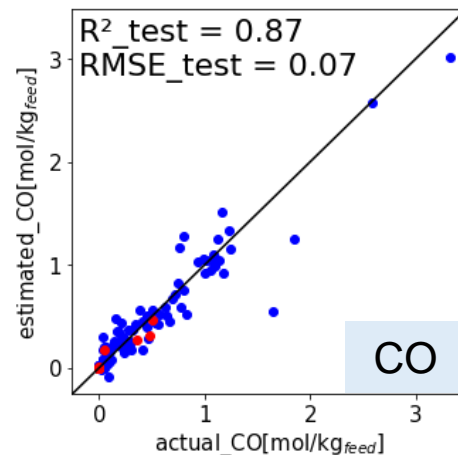
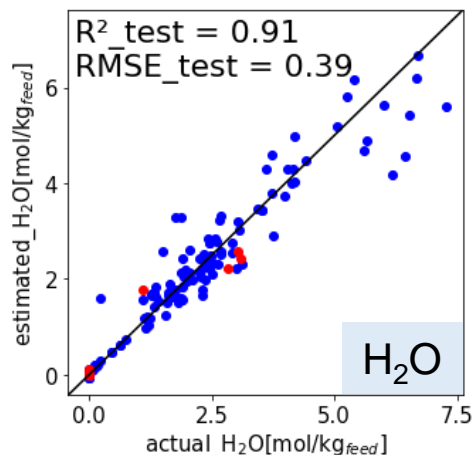
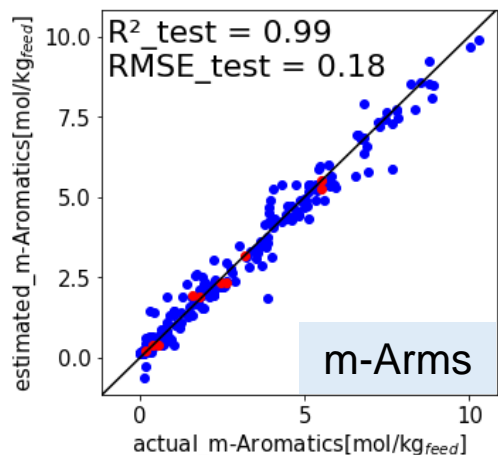
基本特徴量のみ(LASSO) 特徴量の数 = 30



- トレーニングデータ (247点)
- テストデータ (10点)

Physics-based FE model

特徴量の数 = 822



物理化学に基づく特徴量エンジニアリングにより高い予測精度を達成

重要な反応因子の抽出

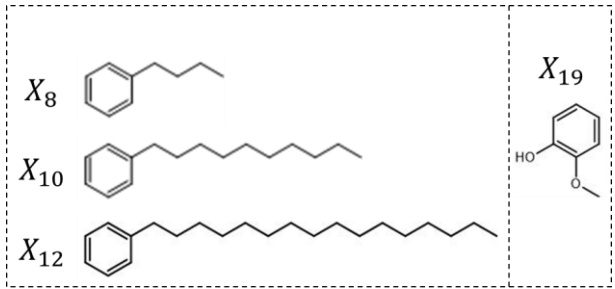
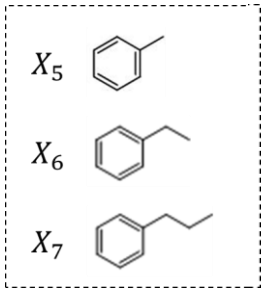
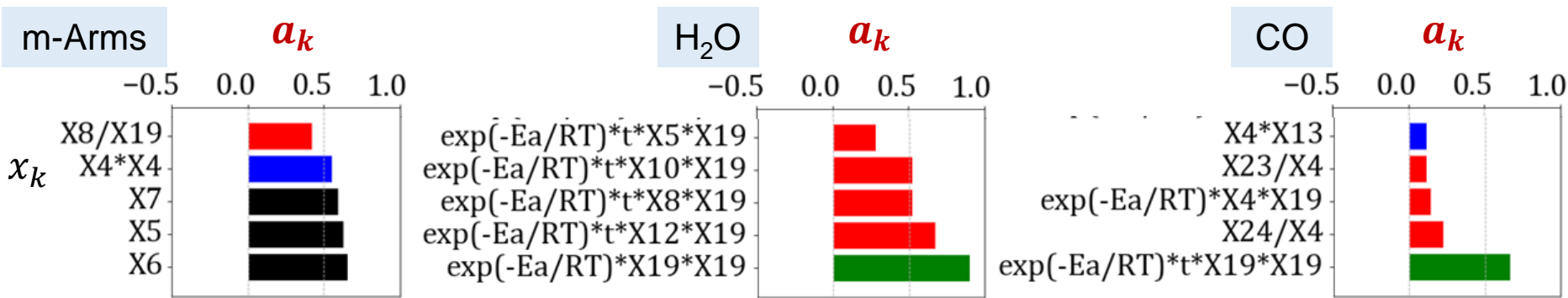
$$y = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 + \dots + a_{822}x_{822}$$

a_k : 標準回帰係数

標準回帰係数の大きさから重要な反応を判断

x_k : 特徴量

* 標準回帰係数の絶対値の大きさ上位5個の特徴量を表示



- 重質油モデル物質の原料濃度
- 相互作用（重質油と重質油）
- バイオオイルの単独反応
- 相互作用（重質油とバイオオイル）

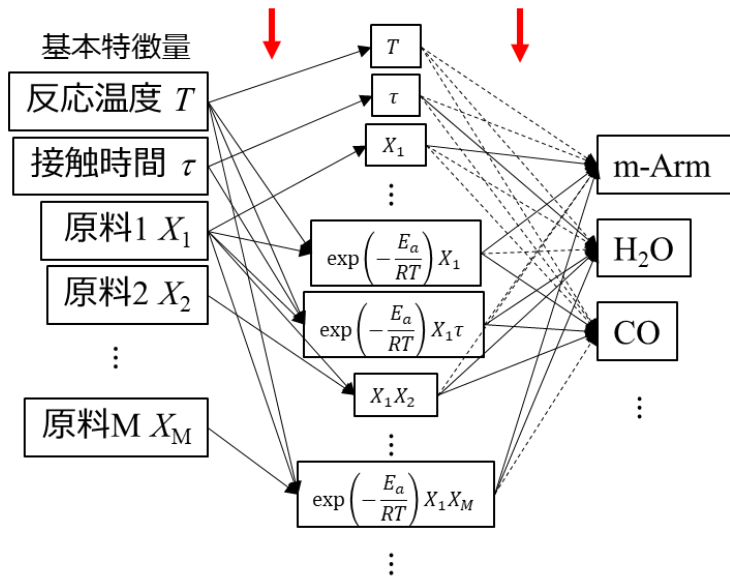
➤ 単環芳香族濃度はアルキルベンゼンの原料濃度の影響が大きい

➤ H₂Oはアルキルベンゼンとグアイアコールの相互作用で生成する

➤ COは相互作用ではなくグアイアコールの単独反応で生成する

重要な反応因子が抽出できる可能性

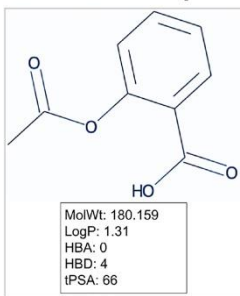
Physics-based FE LASSO



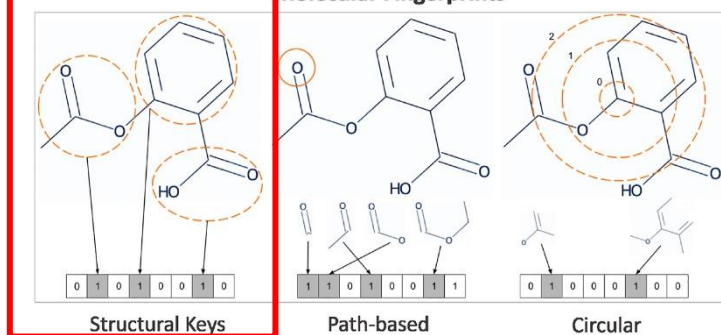
× 未知原料成分の予測が不可能

分子の部分構造の類似性に着目
分子フィンガープリント

Molecular Descriptors



Molecular Fingerprints



Baptista et al., *J. Integr. Bioinform.*, **19**, 36017668 (2022)

分子フィンガープリント
(MACCS fingerprint)

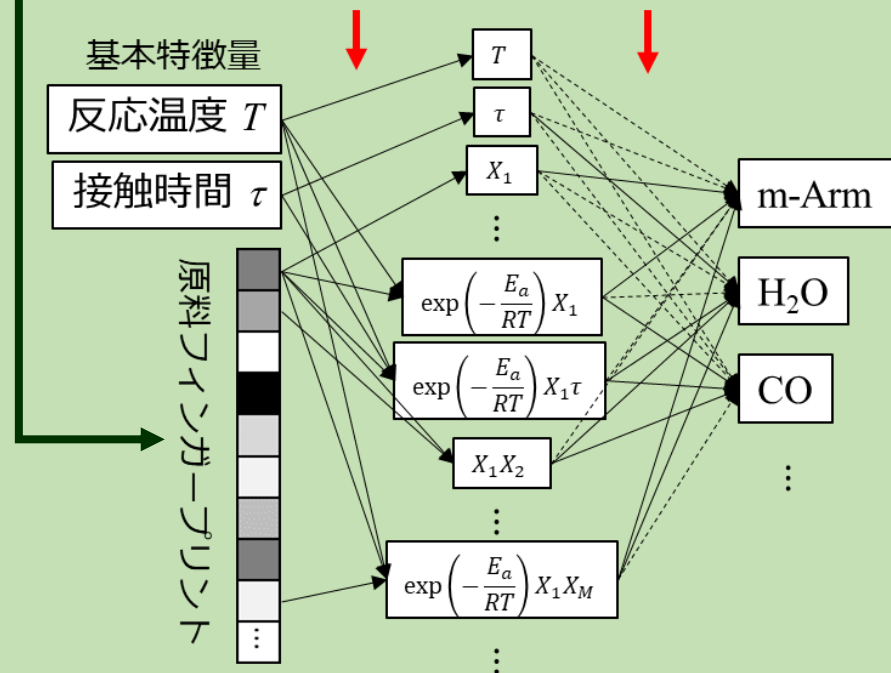
原料1	1	0	0	1	0	0	0	0	1	...
原料2	0	1	0	1	1	0	0	1	0	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
原料M	1	1	0	1	0	0	0	1	0	...

重み付け足し合わせ

原料フィンガープリント

原料1	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	...
原料2	0.0	0.1	0.0	0.1	0.1	0.0	0.0	0.1	0.0	...
原料M	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.1	0.0	...

Physics-based FE LASSO



対象成分を原料に用いた実験データをすべてテストデータとして（訓練データには含まずに）予測精度を評価

● : 対象の原料を含まない実験点 (訓練データ)

➡ モデル構築に使用

● : 対象の原料を含む実験点 (テストデータ)

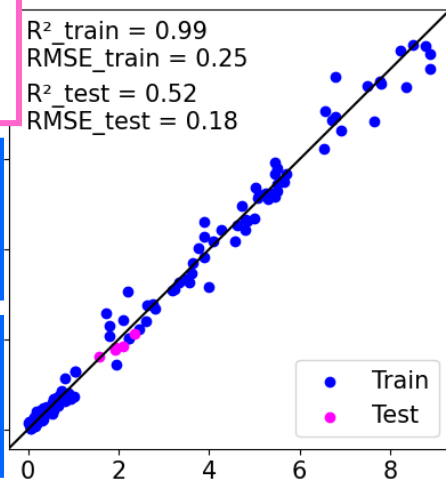
➡ 未知原料を想定した予測

単環芳香族類生成量予測

対象原料 Hexadecylbenzene

重質油モデル物質

$R^2_{\text{train}} = 0.99$
 $\text{RMSE}_{\text{train}} = 0.25$
 $R^2_{\text{test}} = 0.52$
 $\text{RMSE}_{\text{test}} = 0.18$

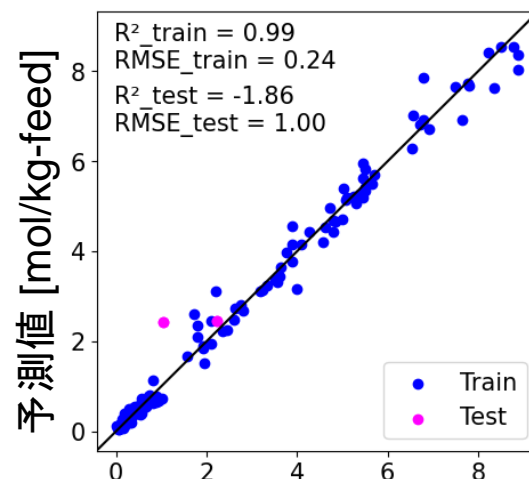


実測値 [mol/kg-feed]

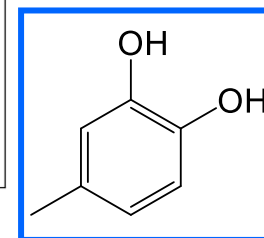
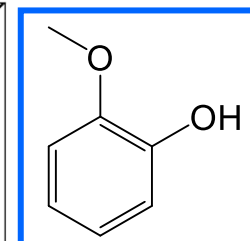
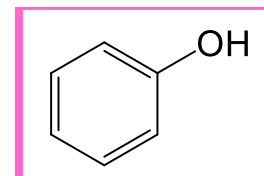
Phenol

バイオオイルモデル物質

$R^2_{\text{train}} = 0.99$
 $\text{RMSE}_{\text{train}} = 0.24$
 $R^2_{\text{test}} = -1.86$
 $\text{RMSE}_{\text{test}} = 1.00$



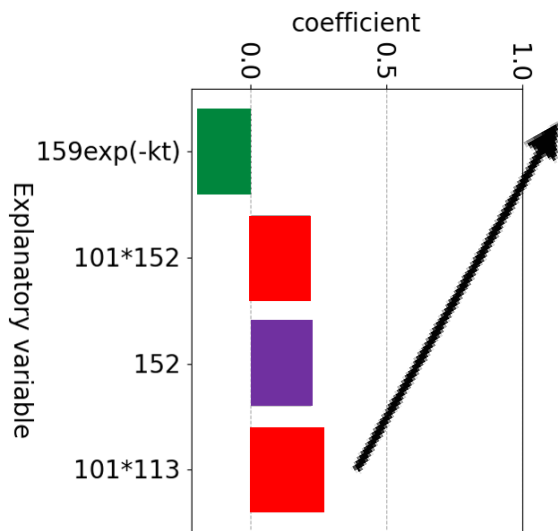
実測値 [mol/kg-feed]



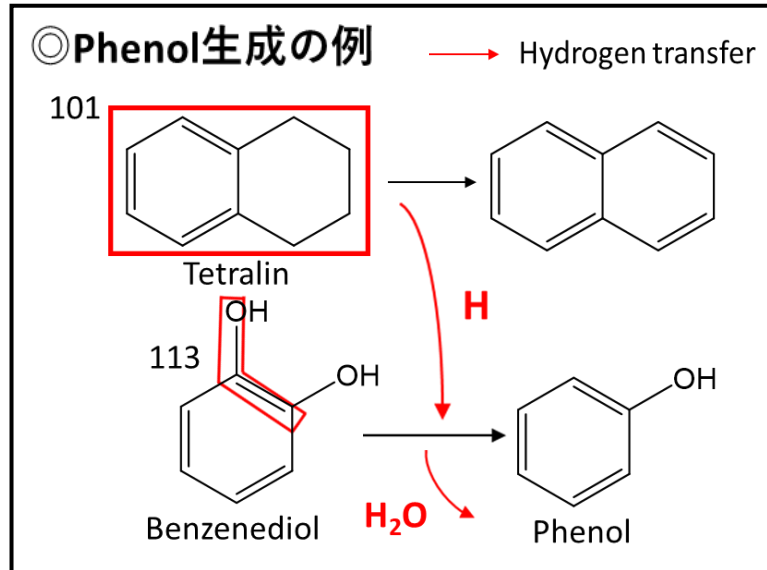
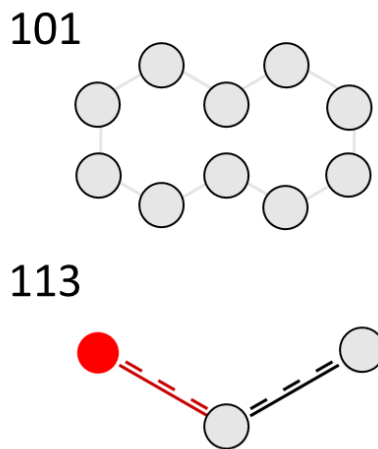
学習に含まれていない原料の反応を高い精度で予測

類似する構造を持つ原料による情報から未知原料の予測が可能

Phenol類生成量予測



・bit 101 × bit 113



- 生成物と類似の構造
- 相互作用 (重質油と重質油)
- 相互作用 (バイオオイルとバイオオイル)
- 相互作用 (重質油とバイオオイル)
- 酸素原子
- 指定無し
- Free bond
- - - Not Aromatic bond
- == Aromatic bond

部分構造によって反応に関する詳細な解釈が可能

まとめ

重質油とバイオオイルの混合接触分解を対象として、機械学習を用いた反応生成物組成予測モデルの構築に取り組んだ。

- 28種のモデル物質を用いた実験データに対して解析を行った
- 物理化学に基づく特徴量エンジニアリングによってLASSO回帰モデルの予測精度が向上することを確認した
- 線形回帰モデルの標準回帰係数から、重要な反応の候補を抽出することができることが示唆された
- 分子フィンガープリントを導入することで、未知の原料成分の反応も予測できる可能性が示された

謝辞

本研究は経済産業省の補助により(一財)石油エネルギー技術センターが実施する技術開発事業の一環として行われたものである。