2023.11.08

一般財団法人
日本建設情報総合センター 助成番号2022-7

下水処理施設における機械学習の利活用に関する予測手法の開発

紀　佳淵

東京大学未来ビジョン研究センター
kikaen@ifi.u-tokyo.ac.jp

東京大学未来ビジョン研究センター
Institute for Future Initiatives
The University of Tokyo

自分について

紀　佳淵（キ　カエン）
Jiayuan Ji (Kaen Ki)

特任講師、東京大学未来ビジョン研究センター
博士（工学、東北大学、2019年9月）
HOPE Fellow、日本学術振興会

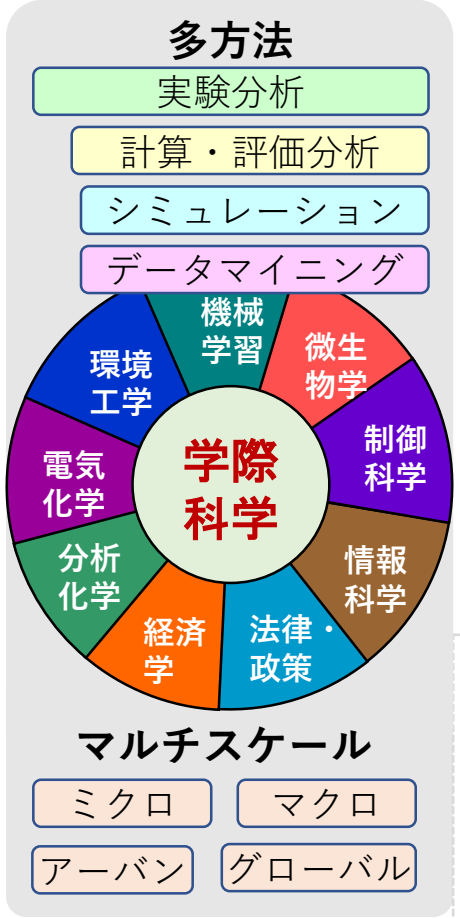
所属学会

日本水環境学会、International Water Association (IWA/国際水協会)、人工知能学会、土木学会、京都大学環境衛生工学研究会、化学工学会、分子科学会、分子シミュレーション学会

編集委員

Frontiers of Environmental Science & Engineering, Sustainability, Electronics, Water…

研究内容

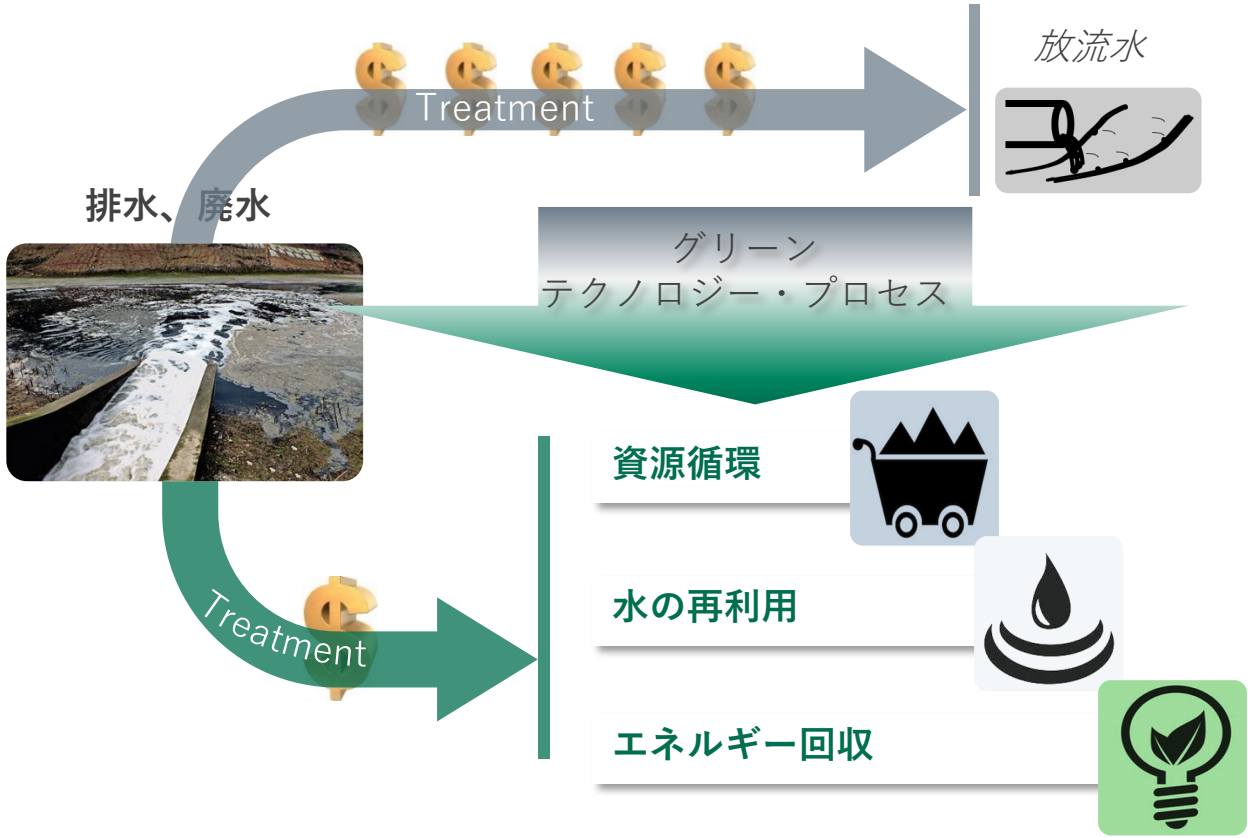


カーボンニュートラル、持続可能な社会

- 廃水・廃棄物処理における資源・エネルギーの回収
- AI/IoTを用いたスマートウォーターシステムの開発
- 新規環境技術の提案・開発・ライフサイクルアセスメント
- シミュレーションを用いたシステムおよび反応装置の最適化

研究背景

- **下水処理施設に対するプロセス革新の方向**



従来法

資金・資源

水の浄化のみ

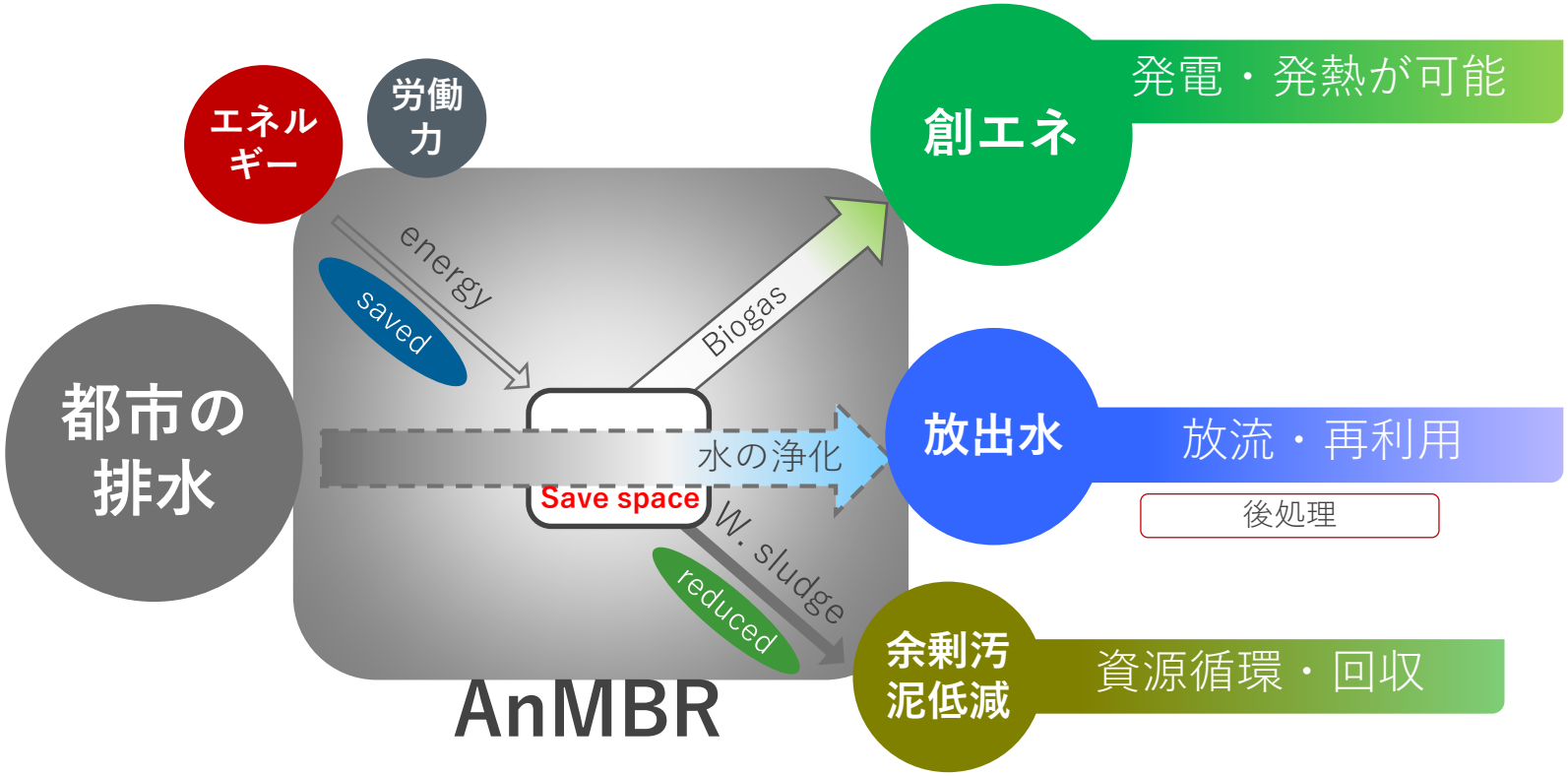
エネルギー消費（処理コスト高額）

温室効果ガス排出（環境・生態破壊）

余剰汚泥など生成（再処理必要）

研究背景

- **嫌気性処理 ＋ 膜分離技術 → AnMBR**



実験装置(過去の研究)

研究背景

- **AnMBRの開発について**

1 実験の方法

経済的コスト:
- 化学試薬
- 専門設備

時間的コスト:
- 数年、数十年ほど

2 伝統的なシミュレーション手段

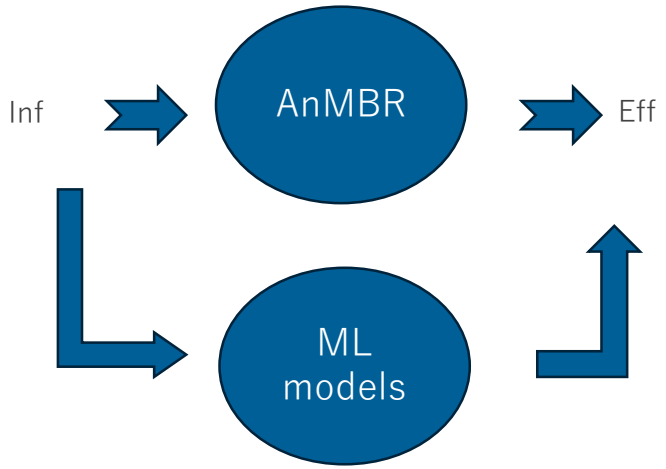
メリット:
- ✓ 時間コストの節約
- ✓ 経済的コストの削減

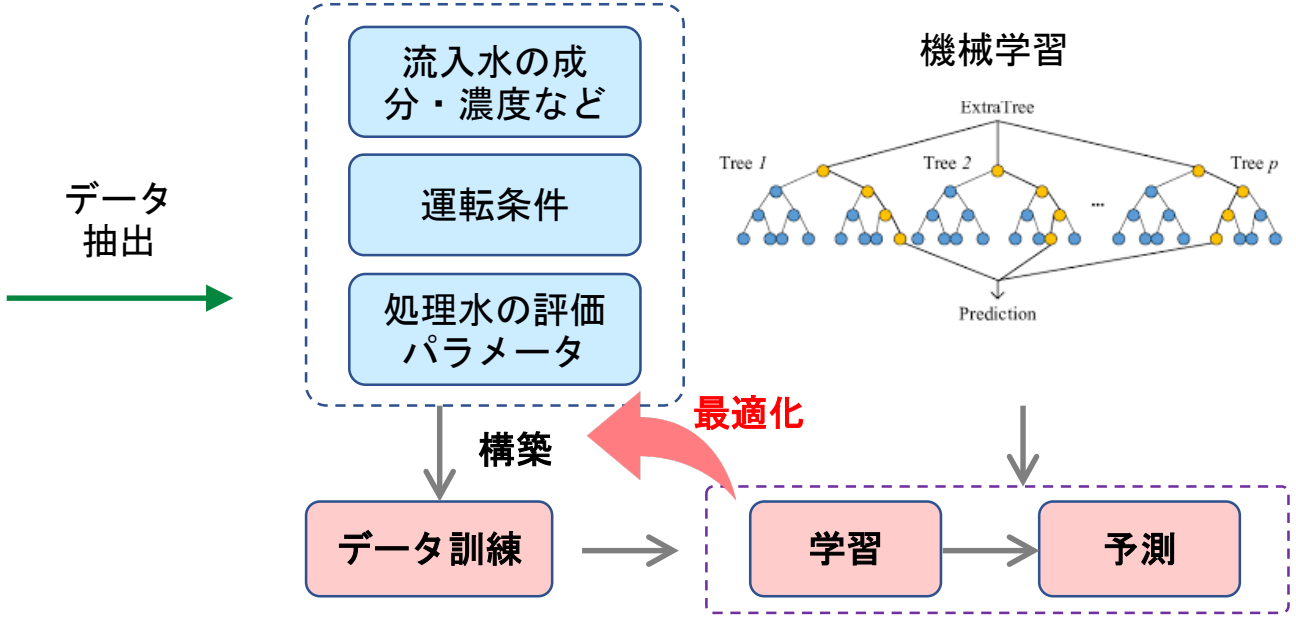
デメリット:
- パラメータが多い
- 信頼性が低い
- 汎用性も低い

機械学習アプローチ

研究方法

- **データセットの構築と予測モデル**



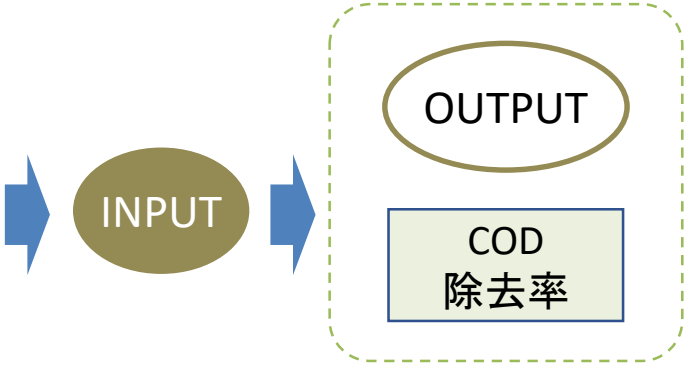


運転条件：水理的滞留時間(HRT)、温度

流入水：pH、水温、有機物濃度(COD)

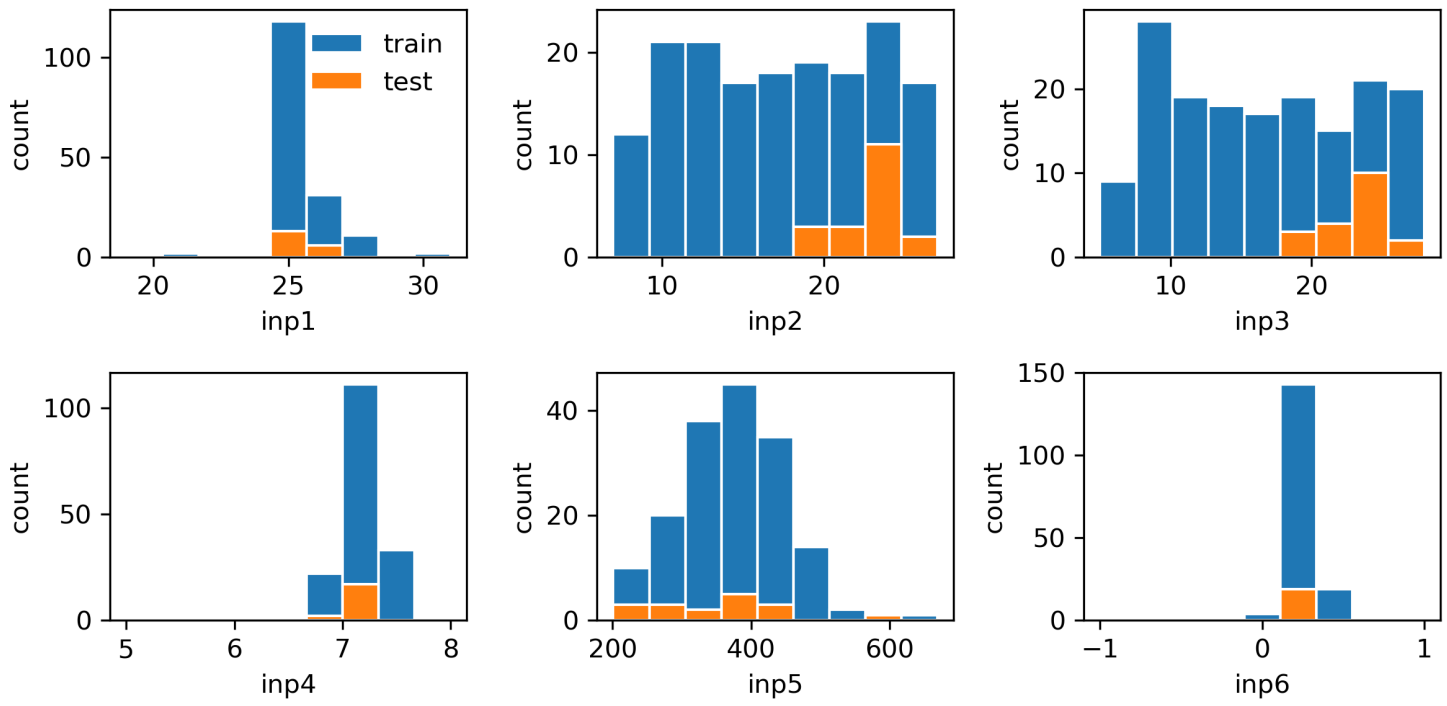
処理水：pH

その他：環境温度、酸化還元電位(ORP)



研究方法

- **データセットの分割とその特徴**



二つのAnMBRにおける一年間の長期運転データにより、トレーニングとテストを9:1でランダムに分割した。

項目	パラメータ
inp0	運転日数
inp1	運転温度
inp2	流入水温度
inp3	環境温度
inp4	流入水pH
inp5	流入水COD濃度
inp6	Flux(流量、処理水量) **水理的滞留時間**を現れる

研究方法

- **ExtraTreesのパラメータ設定**

Tuning:
Hyperparameters: number of estimators, max features, max leaves
Method: 10 folds cross – validation
Criterion: MSE (mean square error)

Model: ExtraTreesRegressor
max features=1.0
max leaves=31
number of estimators=5

- **評価指標**

実測値と予測値による検証：
R^2, MSE, RMSE, MAE, MAPE

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

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2$$

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

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i|$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

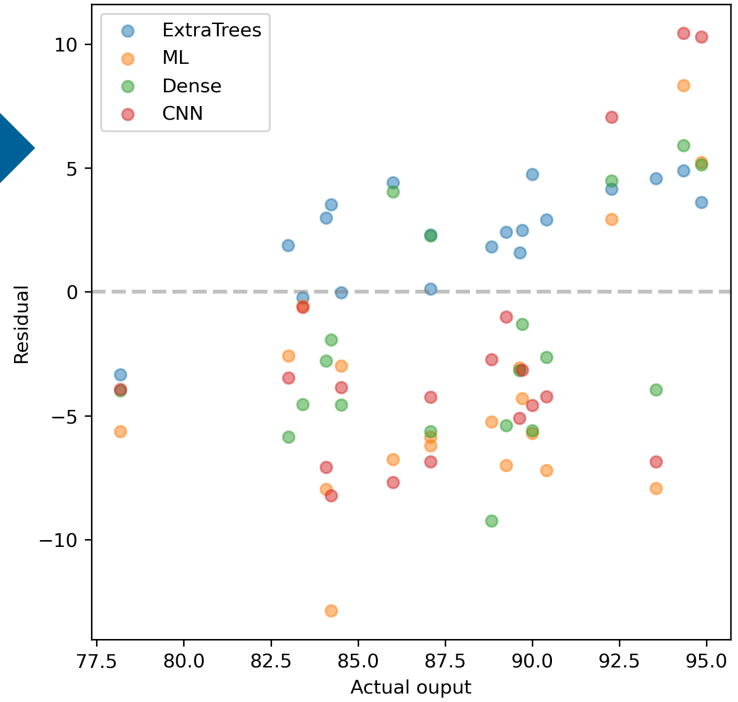
結果・考察

- **OUTPUT = COD除去率に対する予測結果**

Residual = Observation - Prediction

各モデルで得られた結果の比較

RMSE	**3.12** ExtraTrees	6.29 ML	4.69 Dense	5.98 CNN
MAE	**2.74** ExtraTrees	5.26 ML	4.34 Dense	5.34 CNN



ML: machine learning fully connected network
Dense: densely connected convolutional network
RNN: convolutional neural network

参考資料 Science of the Total Environment 813 (2022) 151920

「ExtraTreesは深層学習の三つのモデルに、より良い結果を得られた」

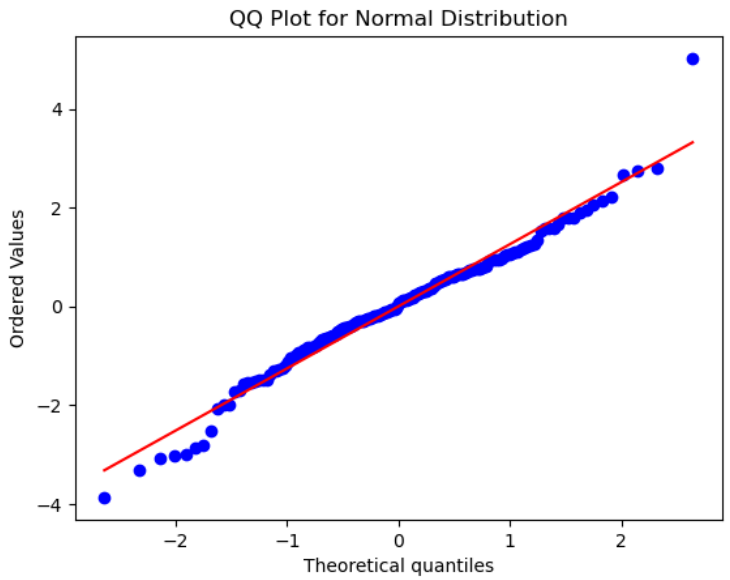
結果・考察

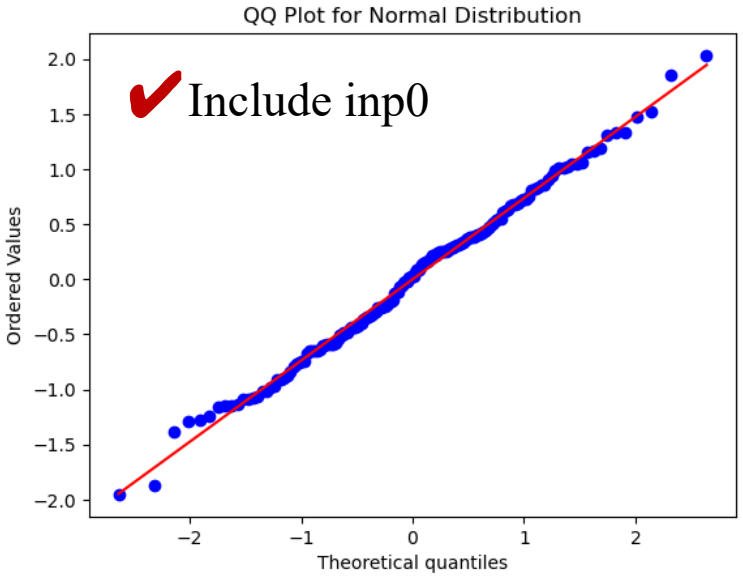
- **予測性能の向上**

①インプットのパラメータ追加により、予測性能を向上させる方法の検討。 そのため、**運転日数(inp0)**をデータセットに追加された。

0.46	9.73	3.12	2.74	0.03
R^2	MSE	RMSE	MAE	MAPE
0.63	**6.62**	**2.57**	**2.12**	**0.02**
R^2	MSE	RMSE	MAE	MAPE

✔ Include inp0

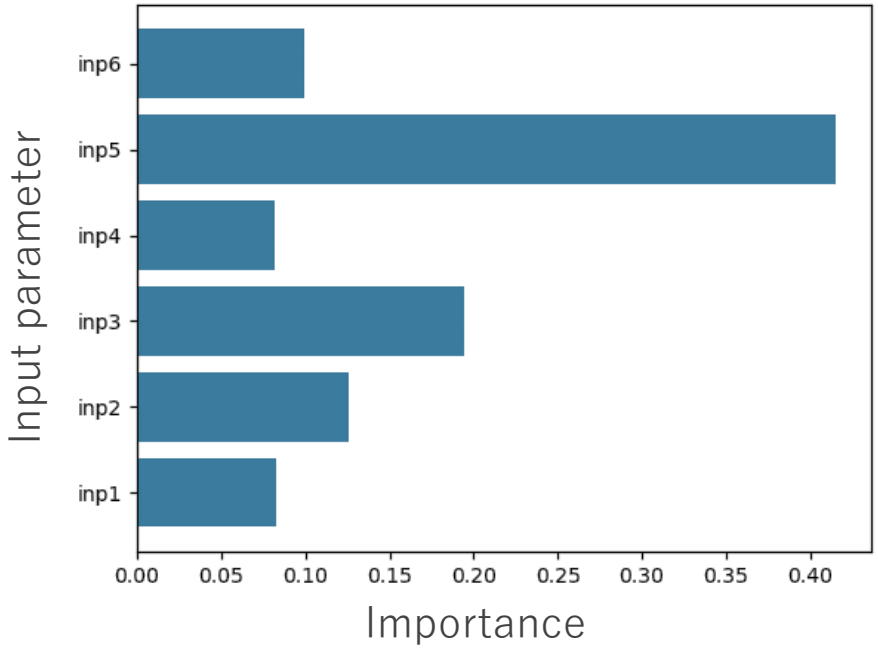


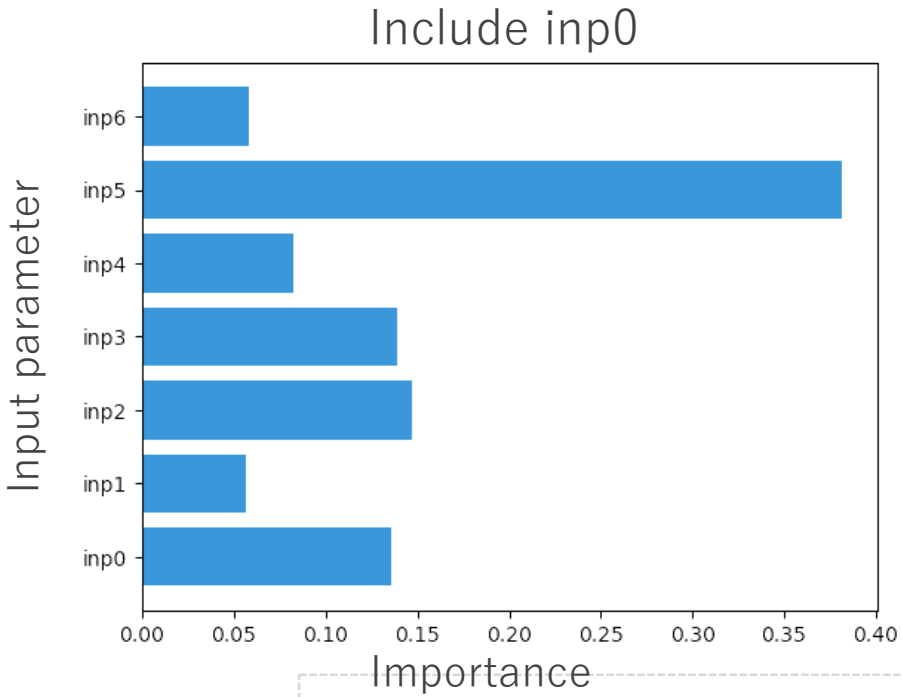


「運転日数(inp0)をデータセットに追加することは、確かに予測性能の向上に効果があることを示唆した」

結果・考察

- **重要度分析**

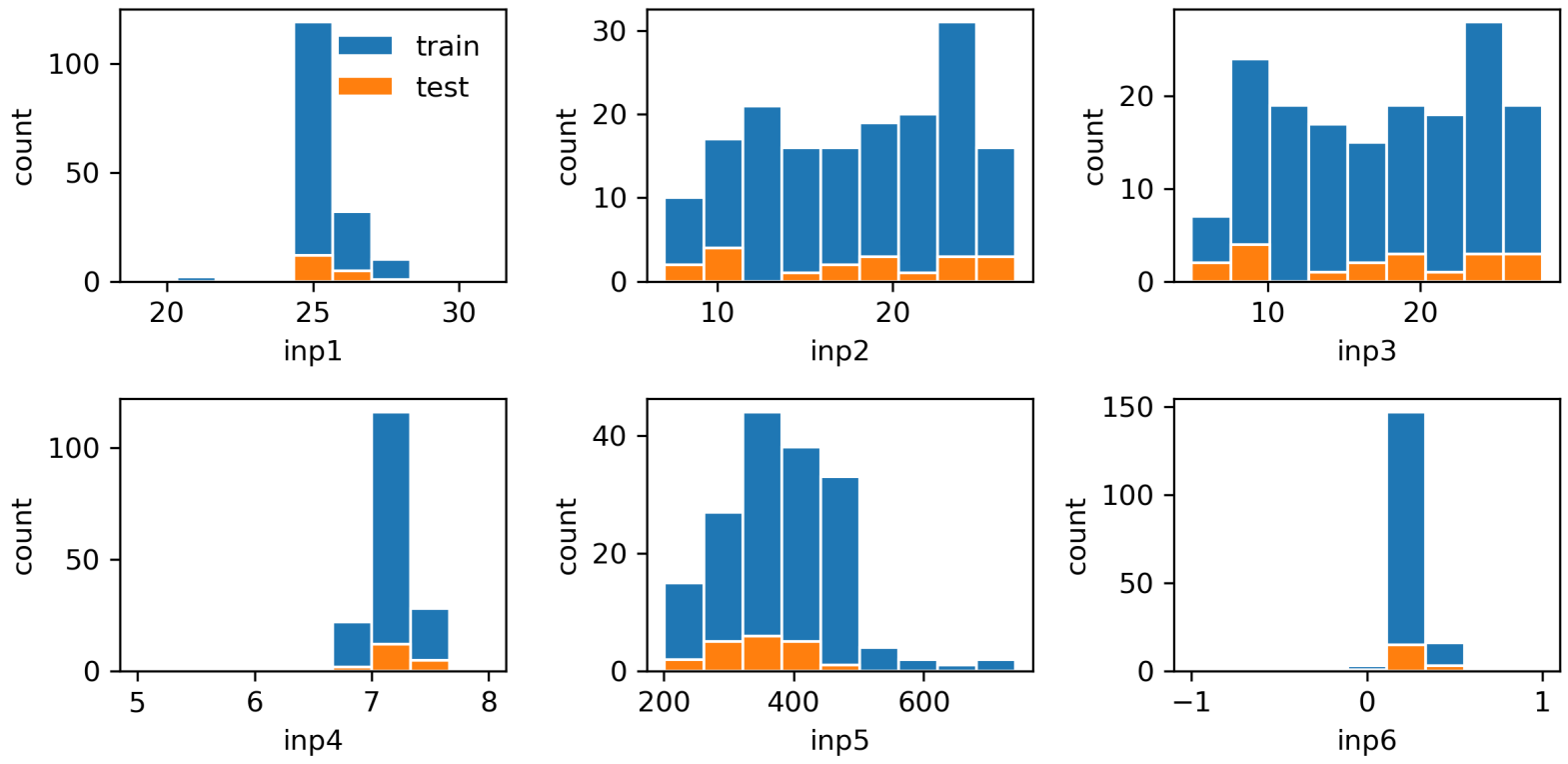




重要度(importance)というパラメーターを求めることで
分類や回帰に寄与した特徴量を把握できる。

結果・考察

- **予測性能の向上**

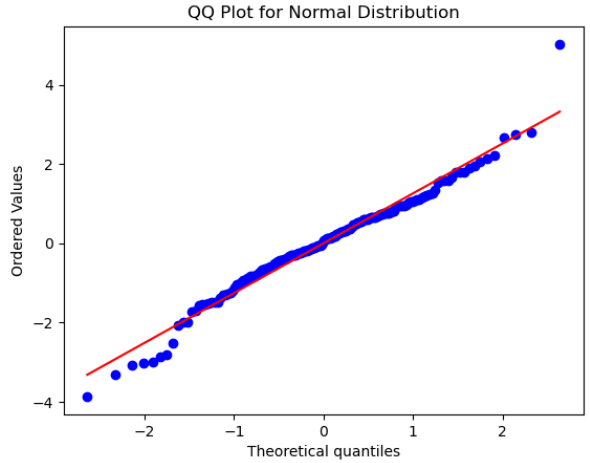


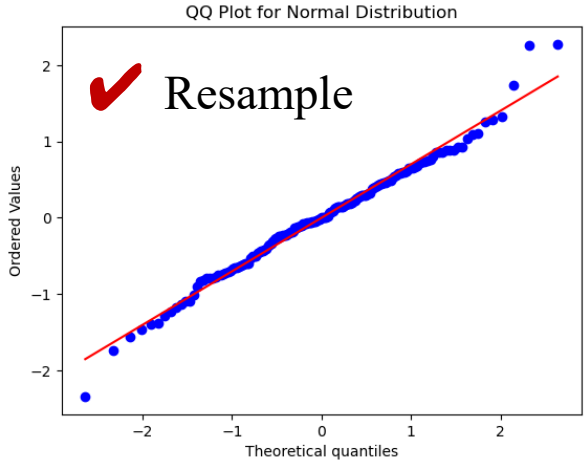
② Resampleにより、予測性能を向上させる方法の検討。

結果・考察

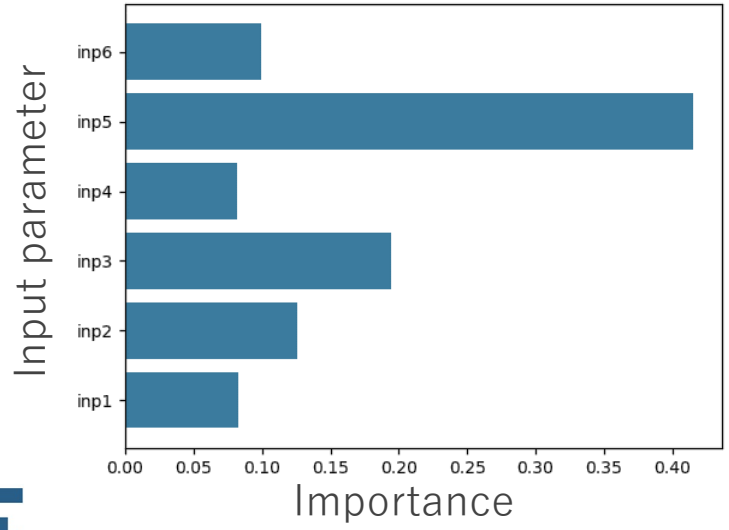
- **得られた結果の分析**

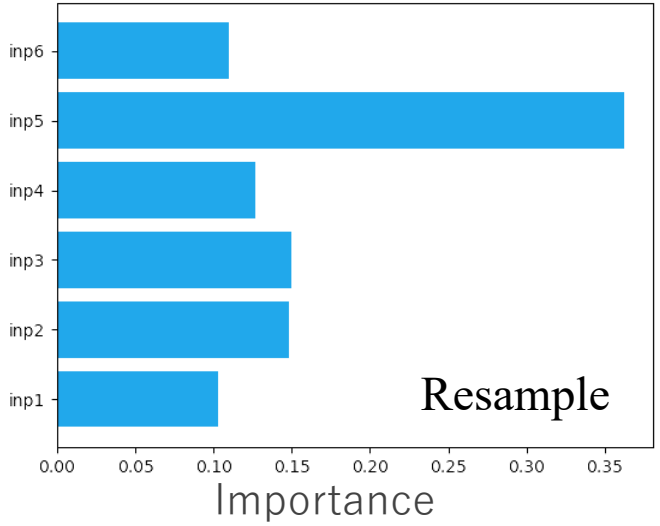
	R^2	MSE	RMSE	MAE	MAPE
	0.46	9.73	3.12	2.74	0.03
✔	**0.66**	**4.69**	**2.16**	**1.89**	**0.02**





- **重要度分析**





「データセットのResampleでは、予測性能の向上に貢献できることを分かった」

結果・考察

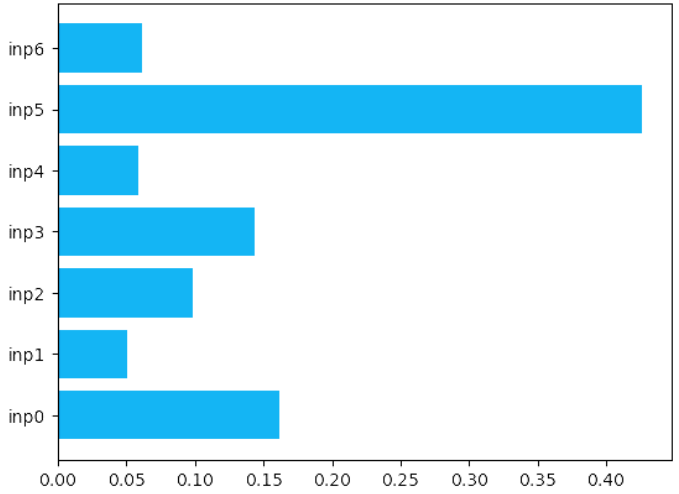
- **予測性能の向上**

③ input0の追加＋Resample

R^2	MSE	RMSE	MAE	MAPE	
0.46	9.73	3.12	2.74	0.03	
0.63	6.62	2.57	2.12	0.02	✔ + input0
0.66	4.69	2.16	1.89	0.02	✔ + Resample
0.73	3.67	1.92	1.63	0.02	✔ + Resample + input0

「input0の追加＋Resampleは、さらに予測性能を向上した」

✔ + Resample + input0



まとめ

- ExtraTreesは深層学習のニューラルネットワークモデルと比較してより良い予測性能を示唆した。
- AnMBRの運転日数パラメータ(inp0)をデータセットの構築に配慮することで、モデリング性能を向上させることが可能である。
- データセットの構築に対して、適切にセグメンテーション(Resample)を行うことで、モデルの効果を確認することが得られる。

今後の課題について

- 予測精度をさらにアップするため、データの高度化およびモデルの選択・開発に関する内容を検討する予定である。
- LSTMという深層学習のニューラルネットワークでは、時系列特徴データに対する予測への適用が良いと考えており、これから進めていきたいところである。

本研究遂行にあたり、研究助成をいただきまして、誠に感謝申し上げます。

今後さらにAI技術の土木環境インフラへの活用のために、研究活動を推進していく所存です。