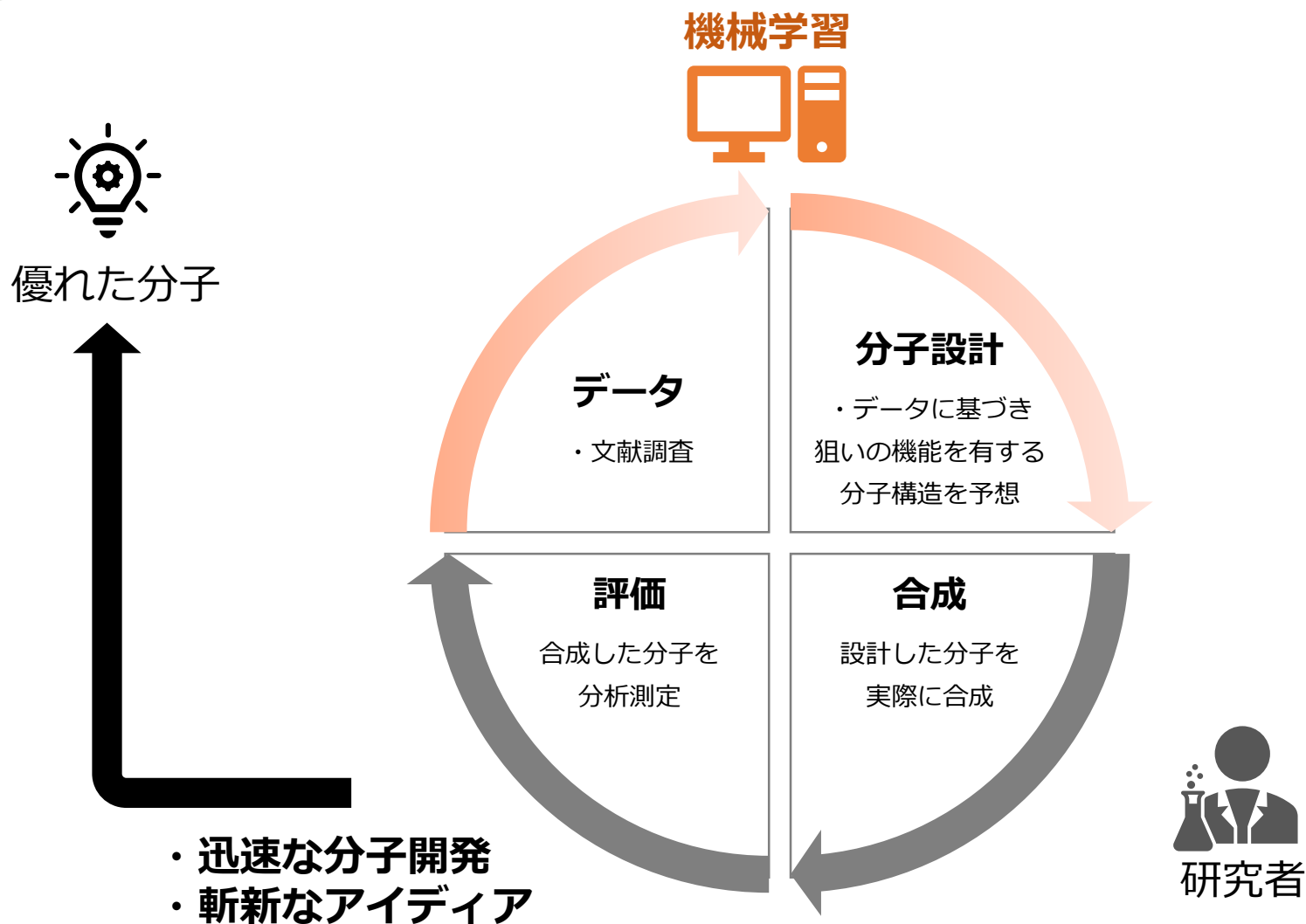


「機械学習による有機電子材料の 物性予測モデルの構築」 に関する報告書

(2021年度 競輪補助事業)



探索プロセスへの機械学習の導入



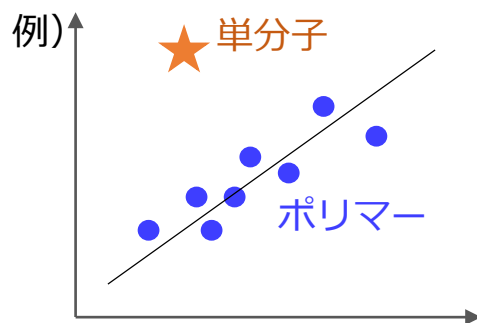
- 多様な学習方法が報告されている
- さまざまな物性値について学習モデル（入力データをもとに作られた型）が作成されている
- よい予測結果が得られている（＝予測値と実測値の一致度が高い）

では、自身が扱う分子群や物性にも適用可能か？

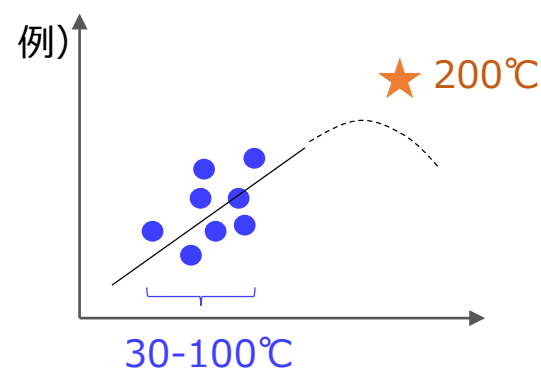


機械学習を利用した報告例とオリジナルの分子

- ・汎用性
(幅広く適用できる性質)



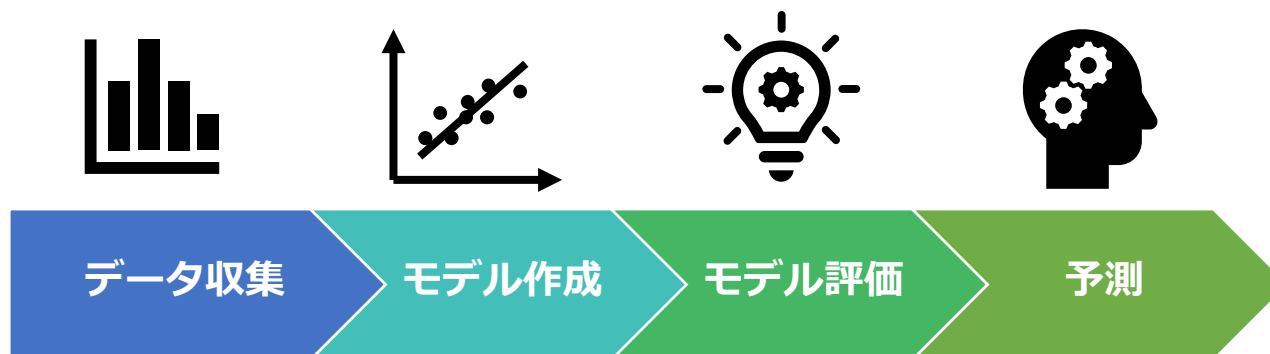
- ・外挿
(データ範囲の外側で予測)



➡ 知りたい分子群や物性に適した各モデルの作成が必要となる

機械学習を用いた予測モデル作成手順

一般的な手順



・プログラミング言語の運用

・データ処理やモデル作成時



統計学が活用

pythonや統計学の習得が必要なため、
予測モデルを作成するには、時間と労力が必要となる

- 機械学習による物性値予測の支援ツールとして、
定量的構造物性相関（QSPR）ソフトウェアを新規導入した
 - 本ソフトウェアを使用して、**予測モデル構築を行った**
 - 1）ガラス転移温度
 - 2）発光量子収率
- 上記モデルを使用して実際に物性値予測を行う
（本事業である共同研究先の和歌山大学大須賀先生の分子で予測）

計算機シミュレーション
に関する知識、情報

和歌山県工業技術センター

化学技術部、地域資源活用部

- ・ 機械学習モデルの構築、検証
- ・ 有用データの収集

有機電子材料に
関する多様な開発実績

和歌山大学システム工学部

大須賀 秀次 准教授

- ・ 化合物、データ選定等に関するアドバイス
- ・ 有機電子材料に関するデータの提供

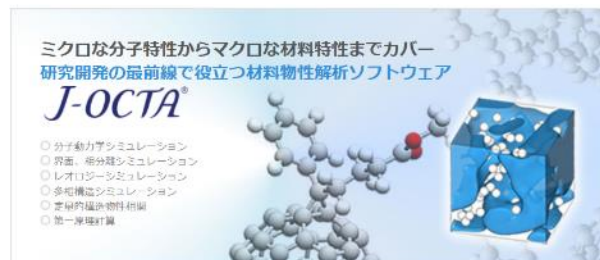


マテリアルズインフォマティクス
による材料開発技術の深耕

ソフトウェア紹介

- J-OCTA（株式会社JSOL）の機能の一つ
- 機械学習による定量的構造物性相関
(Quantitative Structure-Property Relationships, QSPR)

<https://www.jsol-cae.com/product/material/jocta/function/>



◆ ソフトウェアの内容 ◆

- モデル作成～予測までの一般的なステップをソフトウェア上で実施可能
- 学習方法： **グラフ畳み込みネットワーク**などを用いた深層学習
- プログラミング言語のコーディングは**一切不要**

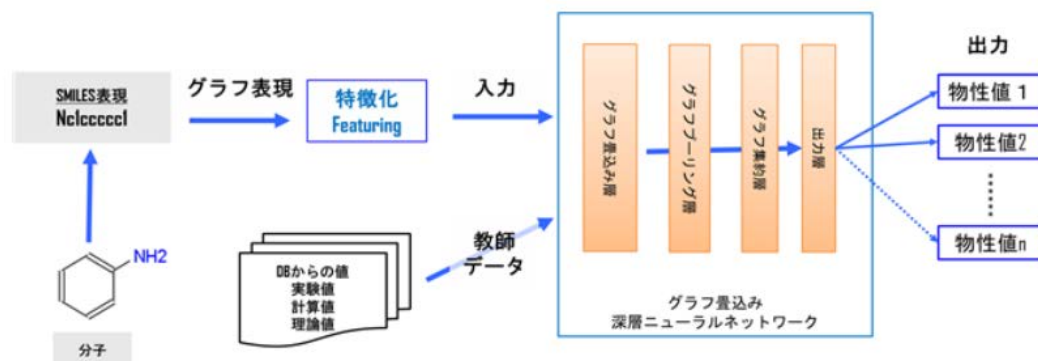


図3 機械学習QSPR

(<https://www.jsol-cae.com/library/function/f20012.html>)

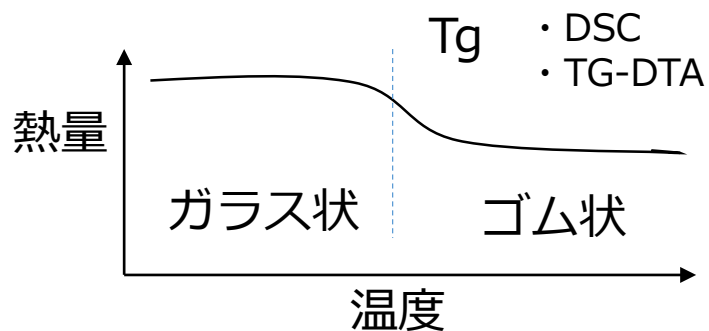
②モデルの評価



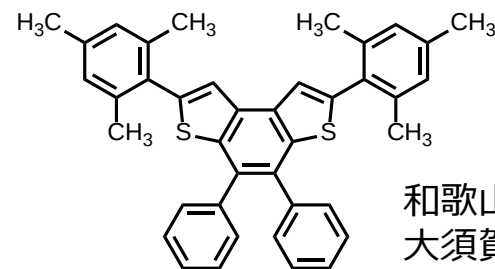
①データの収集

ガラス転移温度に関するモデル作成と物性値予測

ガラス転移温度 (T_g)



【予測対象分子】



Benzodithiophene誘導体 (18分子)

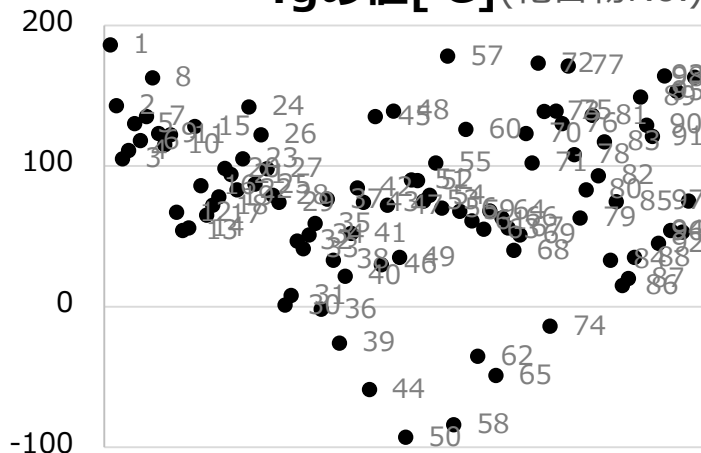
・ T_g はすでに測定済み

◆実験の概要◆

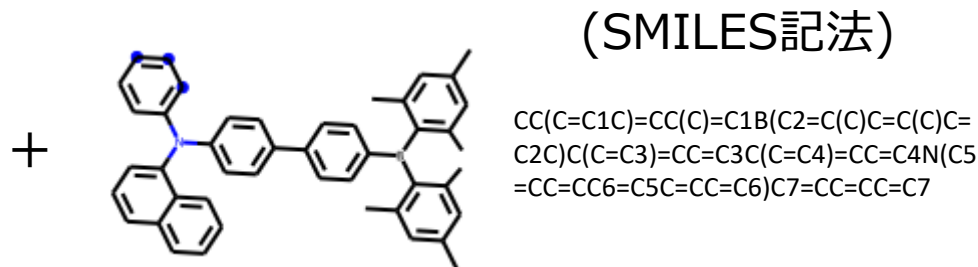


収集したデータを使って学習

Tgの値[°C](化合物No.)



構造式の情報



学習

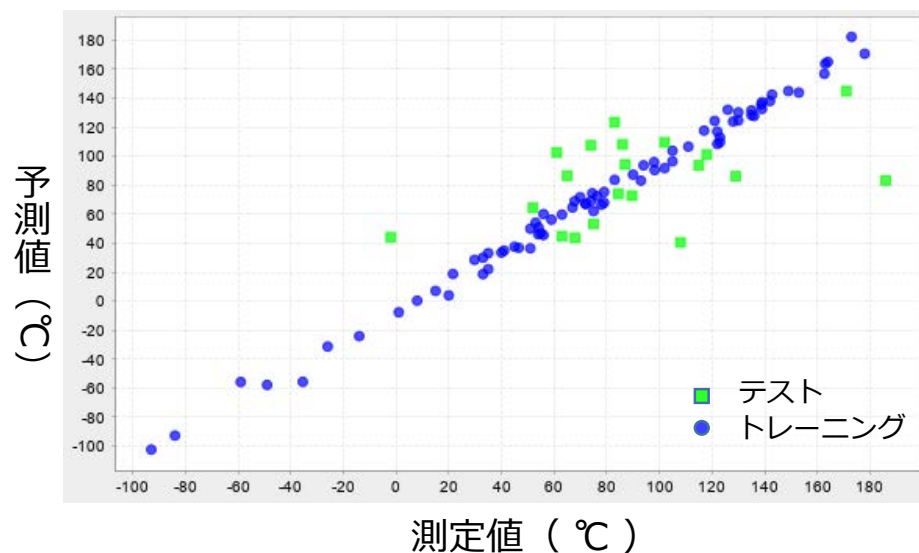
分子構造とTgの値との関係性

データ収集

モデル作成
By ソフトウェア

予測
By ソフトウェア

モデル評価



予測モデル完成！

[モデル作成条件]

Epoc : 15, batch: 10, データ使用率 : 80%

モデルの良し悪しはどう判断するか？

テスト（確認用）データをモデルに当てはめた時の結果（黄緑点）に注目

データ収集

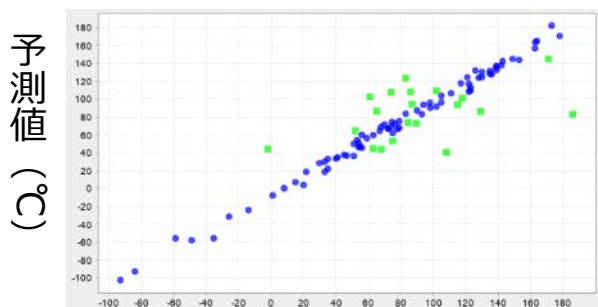
モデル作成
By ソフトウェア

予測
By ソフトウェア

モデル評価

モデルの良し悪しの判断

• 学習結果



■ テスト
● トレーニング

[モデル作成条件]

Epoc : 15, batch: 10, データ使用率 : 80%

	トレーニング (8割のデータ)	テスト (2割のデータ)
MSE	0.018	0.447
R ²	0.984	0.184

MSE (Mean Squared Error、平均平方誤差)
→真値と予測値の差がどれほど大きいかわかる

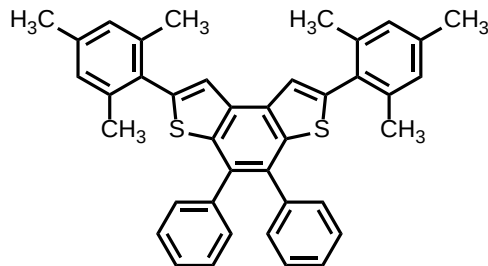
0に近いほど◎

R²値 (決定係数)

→モデルの当てはまりの良さの尺度

1に近いほど◎





予測値と実測値を比較

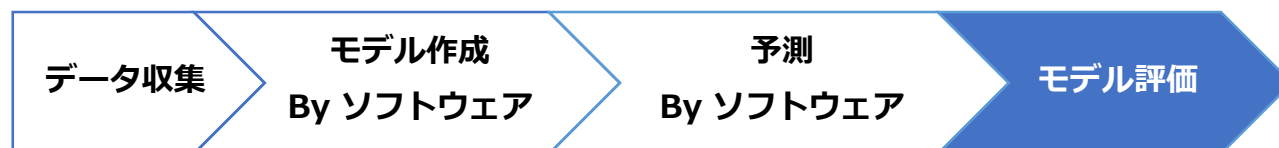
→ 作成したモデルは予測精度が低い

作成したモデル（モデルA）をどう改善するか？

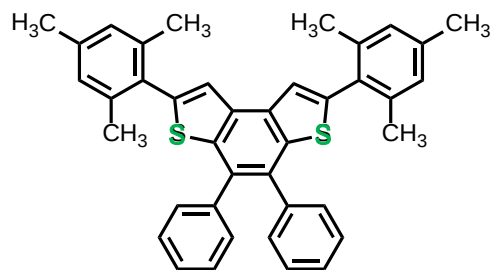
→ 収集データの見直し

検討① データ数を増やす

検討② 類似構造のみを抽出する



モデルCの収集データ



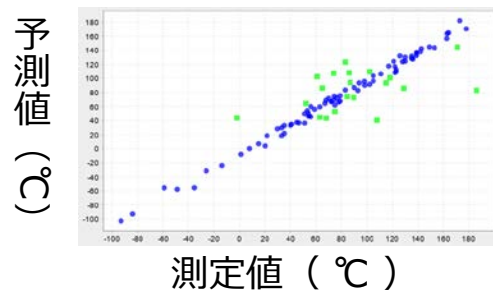
・特徴

1. 芳香族分子
2. 含硫黄

- 同じ特徴を持つ分子構造に限ってデータ収集する

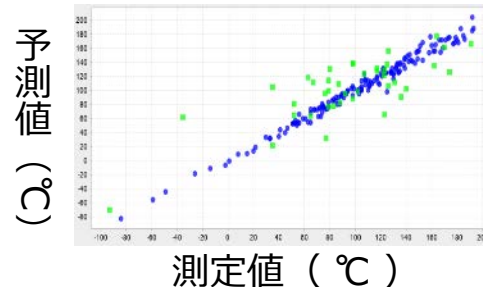
モデルの作成とその学習結果

モデルA



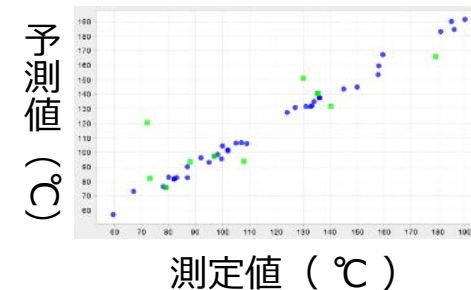
モデルB

(データ数増やす)



モデルC

(類似構造)



■ テスト
● トレーニング

	トレーニング	テスト
MSE	0.018	0.447
R ²	0.984	0.184

	トレーニング	テスト
MSE	0.015	0.417
R ²	0.984	0.593

	トレーニング	テスト
MSE	0.008	0.267
R ²	0.992	0.709

⌈
MSE : 0 に近いほど◎
R²値 : 1 に近いほど◎
⌋

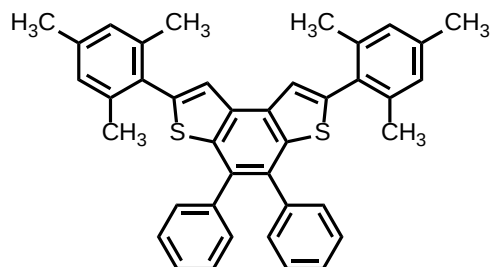
[作成条件]

Epoc : 15, batch: 10, データ使用率 : 80%

- 収集したデータによってモデル（学習結果）が変わる
- MSE,R2値は向上した

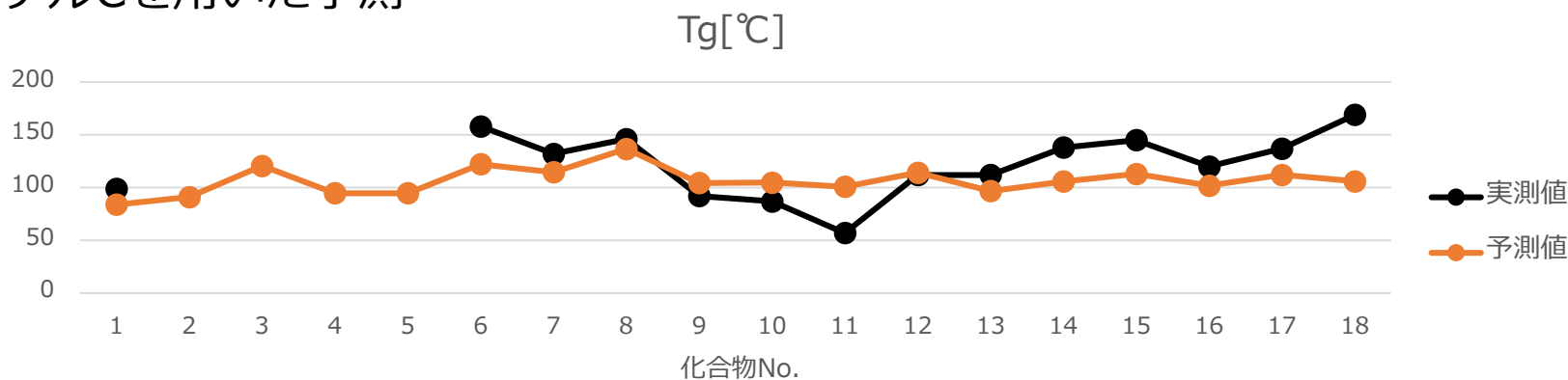
➡ 改善したモデルを使用することで、予測値が実測値に近づいた！

その他の分子における予測とまとめ



+ ベンゾジチオフェン 17件

モデルCを用いた予測



- 差が大きな分子や小さな分子がある
- 精度は高くないが、傾向はつかめている

さらにモデルの汎用性を高めるためには収集データの精査・データの追加による学習の積み重ねが効果的だと考えている

発光量子収率におけるモデル作成と物性値予測

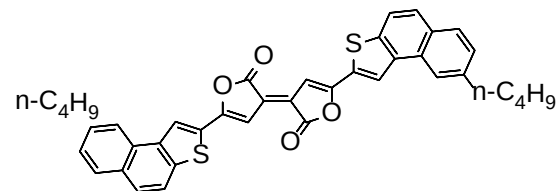
発光量子収率



$$\text{発光量子収率} = \frac{\text{発光フォトン数}}{\text{吸収フォトン数}}$$

例) 有機電子材料として重要な物性値の一つ

【予測対象分子】



+

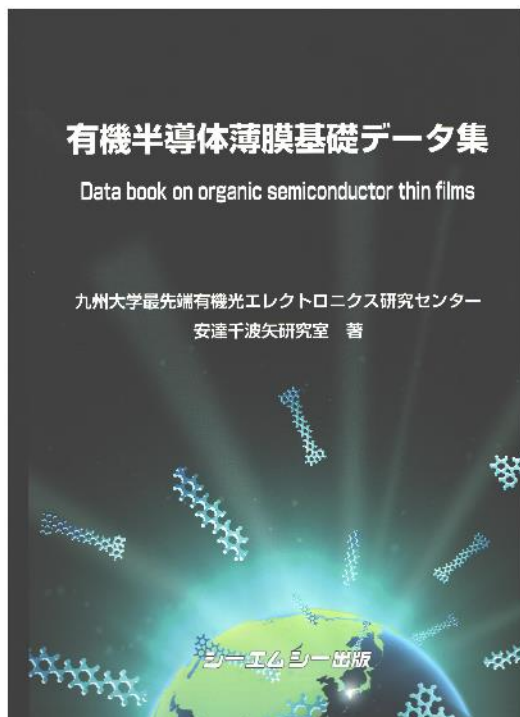
その他の6分子

(和歌山大 大須賀先生から提供)

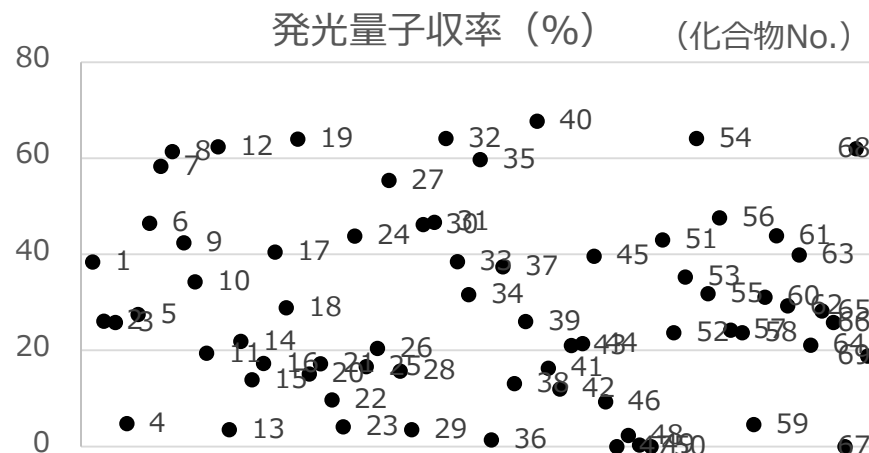
発光量子収率は測定済み

・実験の流れ

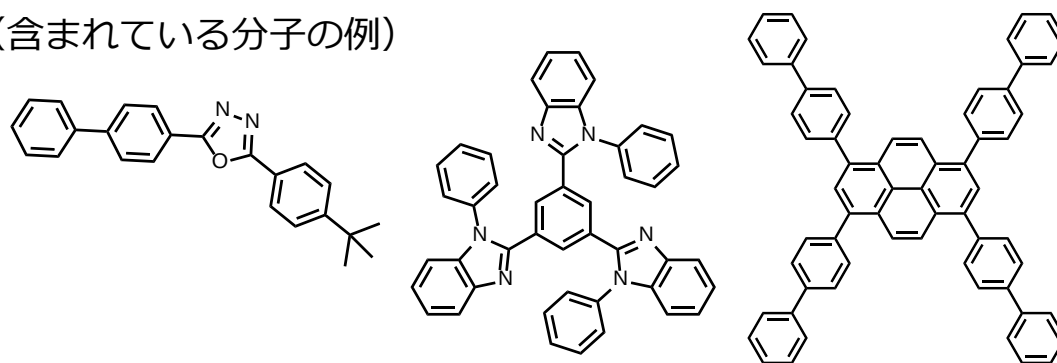




データ : 69件



(含まれている分子の例)



データ収集

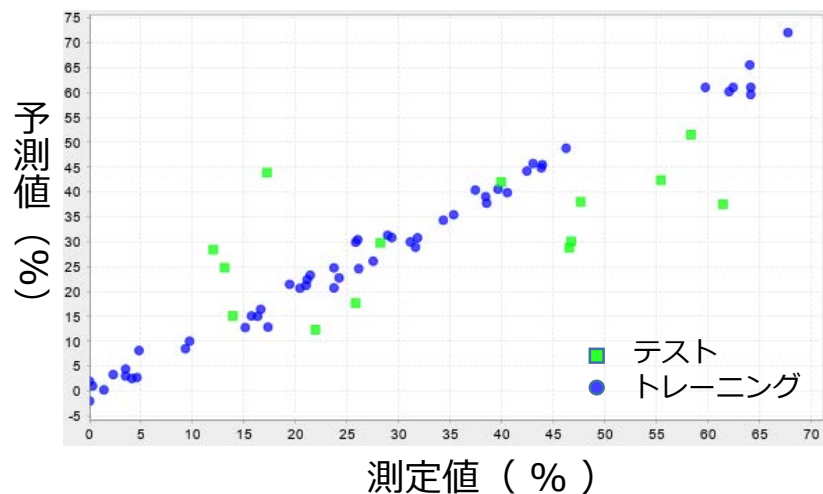
モデル作成
By ソフトウェア

予測
By ソフトウェア

モデル評価

収集したデータでモデル作成

• 学習結果



	トレーニング	テスト
MSE	0.012	0.553
R ²	0.988	0.342

〔 MSE : 0 に近いほど◎
R²値 : 1 に近いほど◎ 〕

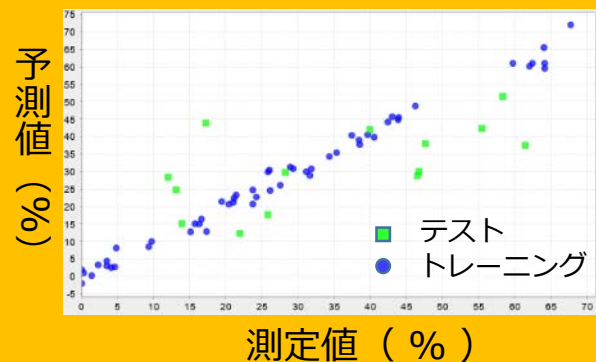
[モデル作成条件]

Epoc : 20, batch: 10, データ使用率 : 80%

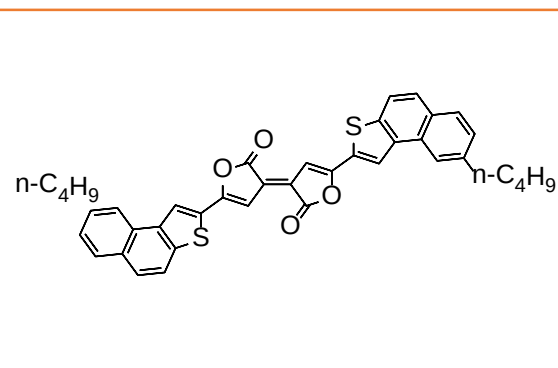


物性値の予測

◆作成したモデル◆



	トレーニング	テスト
MSE	0.012	0.553
R ²	0.988	0.342



数秒後

予測値算出！

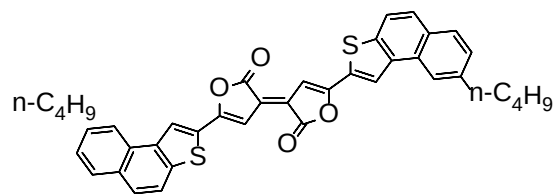
データ収集

モデル作成
By ソフトウェア

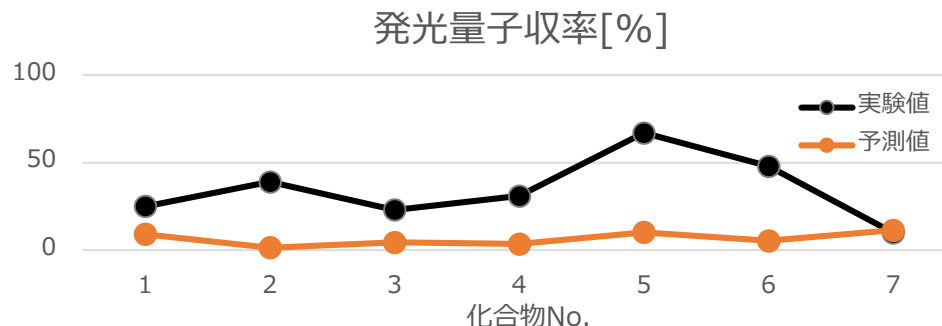
予測
By ソフトウェア

モデル評価

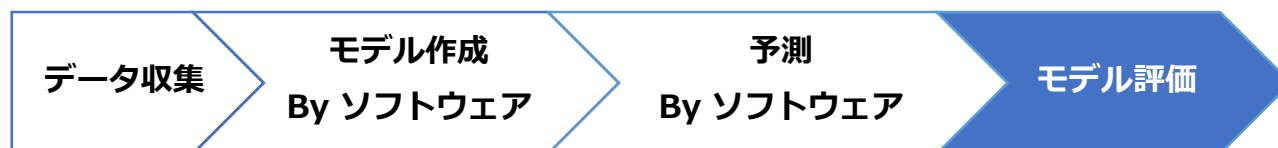
予測値の評価とまとめ



その他の分子も合わせた結果
(合計 7 分子)



- 予測値と実測値の差は分子によって違う
- **収集したデータを見直すことで、同様に改善できる**と考えている
(例えば、データ数を増やす・類似構造の分子群でデータ収集)
- 発光量子収率に関してもモデルを作成できた



ソフトウェアのその他機能

- ▶ 機械学習を行う際に事前に設定するハイパーパラメータの詳細設定、ハイパーパラメータ最適化機能を搭載
- ▶ SMILES取得（PubChem経由やMolファイルから）
- ▶ 各種深層学習機能（グラフ畳み込みネットワークの他、深層ニューラルネットワーク、X G B o o s t）
- ▶ 記述子計算ツール（RDKitに収録されている2D記述子・3D記述子・Fingerprintをソフトウェア上で取得可能）

- 機械学習QSPRソフトウェアを新たに導入
 - モデル作成と予測値を算出（ガラス転移温度・発光量子収率）
 - モデルの改善方法を検討したのち予測精度向上
- 予測したい分子に適応したオリジナルモデルが構築
できることを明らかにした

**ソフトウェアの活用により、
研究開発の迅速化の支援が期待できる**