

適応的実験計画法による マテリアルズ・インフォマティクス

坂口 和優
Kazumasa Sakaguchi

概要

本研究では、ガウス過程回帰を用いて原材料組成からエポキシ系接着剤の物性予測を行った。またその予測結果を用いたベイズ最適化により最適組成設計の自動化に成功したため、その結果を報告する。

1. 緒言

マテリアルズインフォマティクスは材料開発の領域において革新をもたらしつつある。この分野においては、膨大な材料データを機械学習やデータマイニング等の情報技術と結びつけることにより従来のカンコツによる試行錯誤に依存した設計プロセスを著しく効率化している¹⁾。近年、ガウス過程回帰やベイズ最適化といった手法の導入により、材料設計における最適化の自動化が可能となり、望ましい物性を有する材料の組成設計が迅速に遂行されるようになった²⁾。これにより、化学、エネルギー³⁾、医療⁴⁾等多様な分野において、高性能かつ持続可能な材料の開発が促進され、マテリアルズインフォマティクスは次世代の材料開発に不可欠な技術となりつつある。

自動車業界においては、環境規制の厳格化や消費者の環境意識の高まりを背景にEVシフトが促進され、モーターや関連部品のエネルギー効率の向上および軽量化が求められている。中でも電気自動車のモーターに使用される接着剤には厳しい動作環境に耐えうる性能が求められる。高温下でも優れた接着力を保持し、低粘度によって薄膜塗布が可能であることは、エネルギー効率の向上、軽量化において極めて重要な要素といえる。

これらの背景を踏まえ本研究では、モーター向けに開発している高耐熱性低粘度接着剤について、効率的な開発プロセスの確立と、従来の接着剤を上回る性能を持つ組成の探索を目的にマテリアルズインフォマティクスの手法を用いて検討した。

2. 実験

2.1 接着剤

本実験で使用した接着剤は、汎用エポキシ樹脂と高耐熱エポキシ樹脂、4種の希釈剤、2種の硬化剤、6種の硬化促進剤からなる。これら材料を任意の割合で配合した混合物を、10分間脱泡混合し、接着剤を作成した。せん断引張試験はJIS K6848 とJIS K6850に則り試験片を作成し、5mm/minで引張を実施した。粘度測定はJIS Z8803-9に則り単一円筒形回転粘度計を用いて20℃で実施した。

2.2 機械学習モデル

機械学習モデルの構築には、事前の実験で得られた物性値データから、接着強度の指標として20℃雰囲気のせん断引張強度、耐熱性の指標として20℃雰囲気のせん断引張強度を160℃雰囲気のせん断引張強度で割ることで算出した強度保持率、薄膜塗布の指標として粘度の3つを目的変数とした。説明変数は配合組成から14種の原材料を使用した。

回帰モデルにはガウス過程回帰 (Gaussian Process Regression) を使用した。ガウス回帰過程はscikit-learnのgaussian process regression、カーネル関数はScikit-learnのConstantKernel() * RBF() + WhiteKernel()を使用した⁵⁾。ガウス回帰過程の概略図を図1に示す。

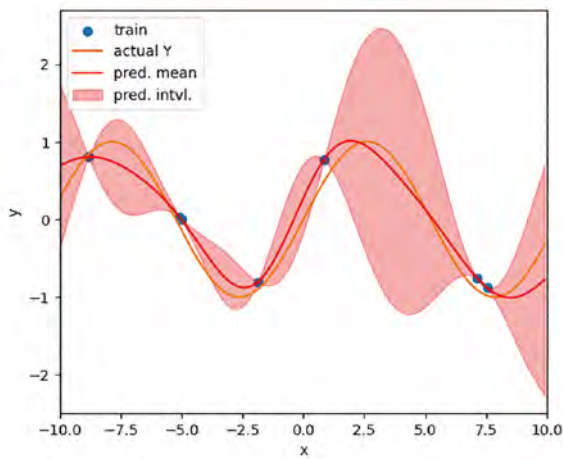


図1 ガウス回帰過程の概略図

青い点が学習データ(実測値),オレンジ線は実際のY,赤線は予測されたY,ピンクの領域は予測されたYの不確かさとして $\pm 3\sigma$ を示す。ガウス過程回帰は説明変数Xと目的変数Yの間にある $Y=f(X)$ を推定する統計手法で,未観測点の予測値と不確かさを同時に出力するため図のように出力Yは推定幅として算出される。データの多い領域では予測区間の幅が狭くなり,データの少ない領域では幅が広がる⁶⁾。

3. ガウス過程回帰モデルの構築

事前の実験で得られた43水準のデータを用いてガウス回帰過程モデルを構築した。データを10分割し,8割を訓練データ,残りの2割を検証データとしてモデルを構築,評価するプロセスを5回行い,その平均値を用いてモデルの精度を確認した(クロスバリデーション, $k=5$)。結果を図2に示す。モデルの性能を評価する指標をグラフ下に記載しており, R^2 (決定係数)はモデルの適合度合を評価する指標,MAE(平均絶対誤差)は予測値と実測値の誤差の平均を取る指標,RMSE(平方根平均二乗誤差)は誤差の二乗平均の平方根を取った指標。 R^2 は1に近いほど,MAEとRMSEは小さいほど優

れたモデルであることを示す。 R^2 が0.5前後の値を取っており,モデルの精度に改善が必要だと考えた⁷⁾。

R^2 を改善するため特徴量の相関解析を行った。相関解析のヒートマップを図3に示す。色が濃いほど高い相関係数を示す。

さらにモデルの精度を改善するため,特に決定係数が低い保持率に対してSHAP値を用いた重要度分析を行った。結果を図4に示す。

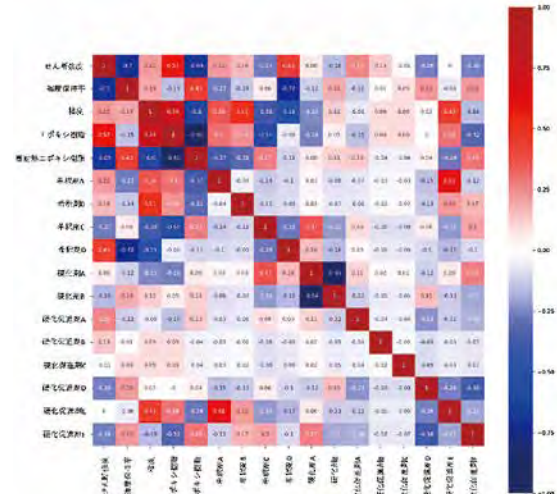


図3 相関係数のヒートマップ



図4 特徴量の重要度(強度保持率)

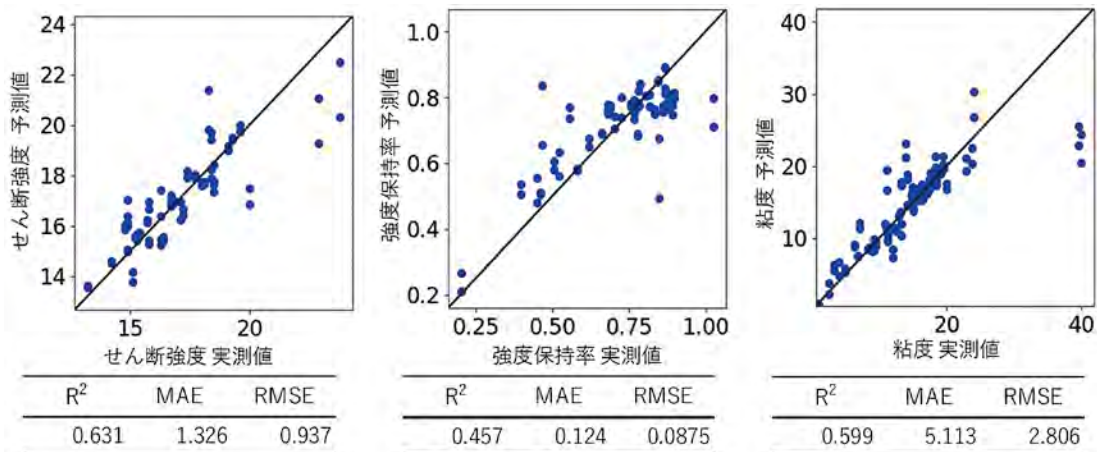


図2 予測精度プロット

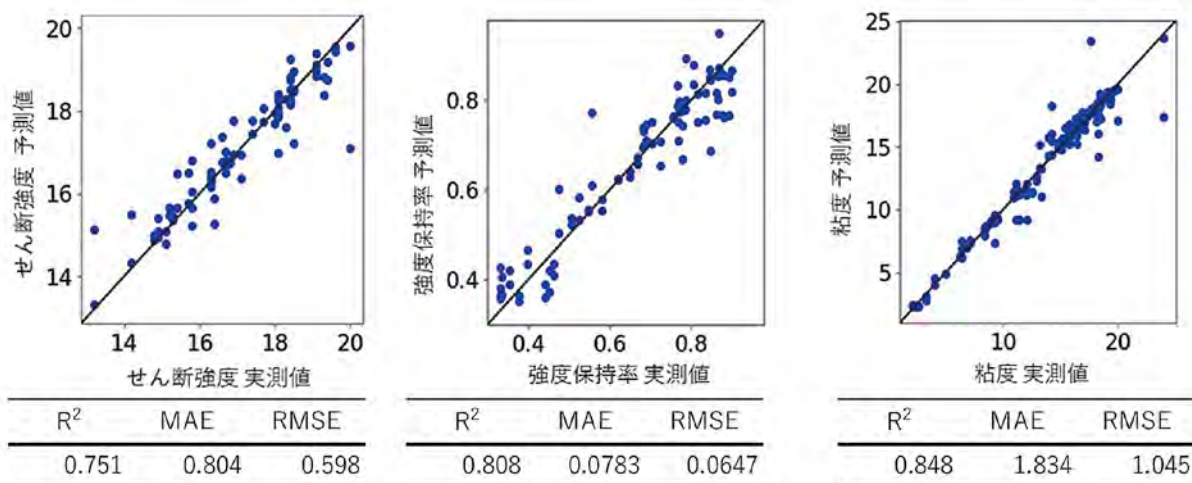


図5 前処理後の予測精度プロット

相関解析が高すぎる項目や、目的変数に対して寄与が小さい特徴量はクラスタリングするか、項目を削除して次元を削減した。また、予測精度プロットで著しい外れ値を示した水準は削除した。

前処理が完了したデータを用いて再度モデルを構築した。結果を図5に示す。全体的に決定係数が向上しており、重要度解析を行った強度保持率ではMAEを大幅に改善することができた。対角線プロットから外れた値も少なくなっており、モデル精度が向上したことがわかる。本研究ではここで作成した3つのモデルを用いて回帰分析を行った。

今回はモデルの予測精度を実際に確認するため追加で実験を行い、モデルが出力した未知な水準の予測値と、実測値の誤差を算出した。結果を表1に示す。

表1 予測値と実測値の誤差

		組成1	組成2	組成3
せん断強度 (MPa)	予測値	17.2	12.4	14.9
	実測値	16.3	11.8	15.1
	誤差	5.3%	5.1%	1.3%
強度保持率	予測値	65.3%	41.3%	51.9%
	実測値	66.9%	43.0%	54.8%
	誤差	2.3%	4.0%	5.3%
粘度 (Pa.s)	予測値	6.1	3.3	8.8
	実測値	6.2	3.6	8.5
	誤差	1.4%	8.3%	3.5%

すべての項目で誤差は数%となっており、実用的な予測モデルの構築に成功したと判断した。

4. 適応的実験計画法

作成したモデルから得られたYの推定幅を利用してベイズ最適化による適応的実験計画法を行った。今回実施した適応的実験計画法は、ガウス過程回帰によりY

とその分散を計算し、得られた値を獲得関数と呼ばれる関数に入力する。獲得関数はYの予測結果が既存のサンプルにおけるYの最大値より大きくなる確率を計算しており、ベイズ最適化はこの値を最大化する点を次の評価候補として選択する。選択された点で実際の評価を行い、モデルに新しくデータを追加する。この流れを繰り返すことで効率よく最適解にたどり着く手法である⁸⁾。今回は獲得関数にProbability of Improvementを使用した⁹⁾。

ランダムに10000水準の組成情報を作成、構築したモデルに入力し、Yの予測値とそのばらつきを算出した。今回は目的変数が3つあるため、YごとにPIを算出し対数変換した値をすべて足し合わせた対数和を獲得関数とした。得られた値を獲得関数に入力し、ベイズ最適化によって、せん断強度と強度保持率が最大化され、粘度が最小化されるXを2つ選択した。獲得関数の値が近いとXも似たような値になってしまうため、初めに獲得関数が最大となるXの水準を1つ選択し、この水準の予測値をYとしてモデルに追加、再度モデルを構築し改めてもう一つの水準を選択することで異なるXを選択しやすくしている。事前に除外していた説明変数はすべて固有の値として特徴量を追加した。

選択された水準を実験しYの実測値を取得、XとYの実測値をモデルに追加し新たなモデルを構築した。このサイクルを3回実施した。結果を表2に示す。

目標値をせん断強度が12MPa以上、強度保持率80%以上、粘度5Pa.s以下と設定した。

1サイクル目では1-1でせん断強度が、1-2で粘度が目標を達成していた。2サイクル目では低粘度を維持しながらせん断強度が向上している。3サイクル目で強度保持率が上昇し始め、6水準目の実験3-2で目標値を達成することができた。

表2 適応的実験計画法の実験結果

		目標値	Cycle1-1	Cycle1-2	Cycle2-1	Cycle2-2	Cycle3-1	Cycle3-2
せん断強度 (MPa)	予測値		16.5	10.2	13.3	15.5	15.9	14.3
	実測値	12以上	16.3	10.8	14.0	15.8	15.7	13.9
	誤差		1.2%	5.6%	5.0%	1.9%	1.3%	2.9%
強度保持率	予測値		69.8%	54.3%	70.7%	70.4%	76.7%	81.2%
	実測値	80%以上	67.0%	59.4%	71.0%	70.8%	75.2%	81.5%
	誤差		4.2%	8.6%	0.4%	0.6%	2.0%	0.4%
粘度 (Pa.s)	予測値		7.0	2.2	5.4	4.4	3	1.7
	実測値	5以下	6.2	2.2	5.5	4.8	2.8	1.7
	誤差		12.9%	0.0%	1.8%	8.3%	7.1%	0.0%

5. 結言

本研究は、ガウス過程回帰を用いて原材料の組成から接着剤の物性を高精度に予測し、その予測結果を基にベイズ最適化による最適組成設計の自動化を行った。これにより、従来の経験則や試行錯誤に依存していた材料設計プロセスを効率化し、迅速かつ的確に望ましい物性を有する材料の組成を探索する手法を提案するものである。

実験結果からは43水準という少ない実験数で複数の物性を同時に高精度で予測できるモデルが構築できることを確認し、ベイズ最適化では各サイクルにおいて設定した目標値を逐次達成し、6水準の実験で全ての目標物性を達成した。この成果は、複合材料の設計におけるベイズ最適化の有効性を示している。

今回実証したマテリアルズインフォマティクスが高性能かつ持続可能な材料の開発ニーズに応えるための手法として今後の材料開発に寄与することを期待する。

筆者



坂口 和優

アイシン化工株式会社
新領域推進センター
接着剤開発に従事

参考文献

- 1) Hatanaka, M (人工知能34): Achievements and Challenges of Materials Informatics in Chemistry, 3-354, 2019
- 2) Wang, J., et al. (School of Operations Research & Information Engineering): Bayesian optimization for materials design, 1, 2015
- 3) Yinbin, X.(表面技術): Bayesian optimization for materials design, 74, 2021
- 4) Oba, F., et al.(J. Am. Chem. Soc.): Band Alignment of Oxides by Learnable Structural-Descriptor-Aided Neural Network and Transfer Learning, 146-9697, 2024
- 5) Pedregosa, F, et al.(Journal of Machine Learning Research, vol.12): Scikit-learn: Machine Learning in Python, 2825, 2011
- 6) Akaho, S.(システム/制御/情報 Vol.62): ガウス過程回帰の基礎, 390, 2018
- 7) Hodson, T(Geoscientific Model Development 15): Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not, 5481, 2022
- 8) Kaneko, H.(講談社サイエンティフィク): Pythonで学ぶ実験計画法入門, 116, 2021
- 9) R.P.Adams(Journal of Machine Learning Research): Practical Bayesian Optimization of Machine Learning Algorithms, 1, 2012