

マテリアルズインフォマティクスを活用した材料技術開発

基盤技術センター 島田 真至¹・川本 慎二²・木 崙 剛志³

Materials Technology Development Using Materials Informatics

M. Hatada, S. Kawamoto, and T. Kizaki

従来の実験科学や理論科学に加え、計算科学やデータ科学などの計算機科学を材料開発に積極的に活用するアプローチのことをマテリアルズインフォマティクス (MI) と呼ぶ。当社では、MIを基盤材料技術の1つと位置付け、多種多様な製品への適応を図ってきた結果、様々な成果に結びついている。本稿では、使用頻度が高いMI手法を紹介した後に、ケーブルのシース開発にデータ科学を活用することで、開発期間の大幅な短縮に成功した事例を紹介する。

Materials informatics (MI) is an approach that actively utilizes computer science, such as computational science and data science, in materials development in addition to conventional experimental and theoretical science. We have positioned MI as one of the fundamental material technologies and have applied it to a wide variety of products, which has led to various outcomes. In this paper, after introducing MI methods that are frequently used, we introduce a case in which we succeeded in significantly shortening the development period by utilizing data science in the development of cable sheaths.

1. ま え が き

近年、材料分野においてマテリアルズインフォマティクス (MI) が注目されている。2012年10月に、MITと韓国サムスンがイオン伝導度が非常に高い固体電解質を開発したと発表した¹⁾。しかし、それは日本の大手自動車メーカーが以前より数年かけて取り組み、2011年に特許出願した物質とほぼ同じだった。さらに、この開発は一切実験しておらず、データと計算を利用した手法のみで開発を行ったことから、世界中の材料研究者にとって非常に大きなインパクトとなった。このことをきっかけに世界中でMIが潮流となり、次々と国家プロジェクトが立ち上がった。日本では2013年に国立研究開発法人 科学技術振興機構 (JST) がMIの促進を提唱²⁾して以来、3府省が国家プロジェクトを開始した。現在では、MIは材料技術開発を支える基盤技術の1つであると認識されはじめており、様々な企業で本格導入が進んでいる。

当社は事業領域が幅広いことから多種多様な材料を扱っており、製品の競争力強化につながる基盤材料技術の強化は必須であることから、2018年からMIへの取り組みを進めてきた。MIはまだ発展途上であり、正式な定義は存在しないが、当社では新規材料開発のみではなく、プロセスの最適化なども含めてMIと定義している。また、計算機

科学はあくまでツールであり、MIというツールを使いこなすためにはドメイン知識 (MIを適用する領域での専門知識、知見、経験、ノウハウ) が重要であることがわかってきており、材料エンジニア自身がデータサイエンスを学ぶための機会の提供や計算環境の構築などを推進することで、より実用的な成果に結びつける活動を展開してきた。その結果、ケーブルのシース開発において、配合最適化の実験条件立案にデータ科学を活用することで、大幅な開発期間の短縮に成功した。

2. 主なMI手法の概要

材料技術開発ではMI手法の中でも、「ベイズ最適化」、「回帰モデル、分類モデルの作製」、「モデルの解釈」、「パーシステントホモロジー」などの活用頻度が高い。この4つの手法の概要を紹介する。後述のケーブルシース開発において、使用した主なMI手法はベイズ最適化とモデルの解釈の2つである。

2.1 ベイズ最適化

MIでは、ベイズ最適化が活用されることが多い³⁾。ベイズ最適化とは、今までの実験データから次に実験する価値が高い条件を導きだし、逐次的に実験計画を立案しながら最適化していく手法である⁴⁾。図1にベイズ最適化のイメージ図を示す。図1 (a) ~ (c) において、横軸はパラメータ、縦軸は目的変数であり、青丸のプロット点は実験結果であ

1 材料技術開発部

2 材料技術開発部 主査

3 材料技術開発部 グループ長

略語・専門用語リスト

略語・専門用語	正式表記	説明
MI	マテリアルズインフォマティクス	材料開発に計算科学やデータ科学を積極的に活用する手法。
回帰モデル		学習データに正解を与えた状態で機械学習を実施する教師あり学習の一種で、正解が数値データであるもの。
分類モデル		学習データに正解を与えた状態で機械学習を実施する教師あり学習の一種で、正解がカテゴリデータであるもの。
OFAT	One Factor At a Time	対象とする説明変数以外は固定して、一つの説明変数のみが目的変数に与える影響を調査する手法。一時一事法とも呼ばれる。
D最適計画		Xの行列式 $\det((X'X)^{-1})$ が最小になるように実験計画をすることで、交絡を少なくする手法。
UMAP	Uniform Manifold Approximation and Projection	次元削減手法の1つ。比較的高速な処理が可能であり、高次元データの可視化によく利用される。
Permutation importance		説明変数をシャッフルしたときの予測誤差の変化を利用して説明変数の重要度を評価する手法。順列重要度とも呼ばれる。
Partial dependence plot		着目する説明変数以外の影響を周辺化により除去することで、一つの説明変数と目的変数の関係を解釈する手法。部分依存プロットとも呼ばれる。

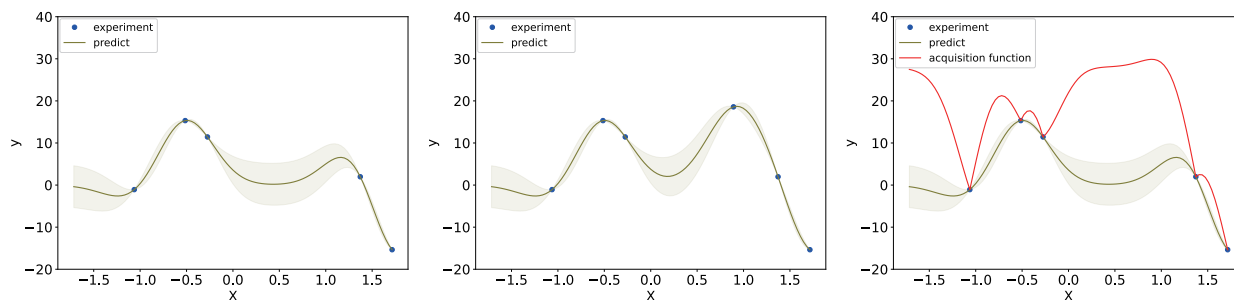


図1 ベイズ最適化の概要

(a)確率モデル作製, (b)獲得関数算出, (c)データの取得と確率モデルの更新

Fig. 1. Summary of bayesian optimization.

(a) Create probability model, (b) Calculate the acquisition functions, (c) Experiments and updating of probability model.

る。ベイズ最適化ではまず、実験データから確率モデルを作製し、探索範囲における平均値と分散を推定する。その結果を図1 (a) に示す。青線が平均値、灰色の領域が分散を示している。この結果を基に実験する価値を表す獲得関数と呼ばれる指標を算出する。図1 (b) の赤線が獲得関数を示している。この獲得関数が最も高い条件で次の実験を行い、その結果を追加し確率モデルを更新させた結果が図1 (c) となる。これらを繰り返すことで最適な条件を探索していくのがベイズ最適化である。

従来は経験と勘から当たり付けをし、一つの変数のみを変更させた際の目的変数の変化を把握するOFAT (One Factor At a Time) と呼ばれる手法を用いながら、絨毯爆撃式に実験を繰り返すことで、各パラメータと目的変数

の関係を明らかにしていくことが多かった。しかし、この手法では局所的な探索になってしまう傾向があるという課題があった。それに対し、ベイズ最適化では獲得関数を適切に設定することで比較的大域的な探索を実施することが可能である。

また、ベイズ最適化において確率モデルを作製するには初期データが必要である。初期データ収集時の実験計画立案方法としては、直交表を活用する方法やD最適計画を活用する方法などが提案されている^{5) 6)}。なお、直交表では水準数を固定するのに対し、D最適計画では水準数を固定する必要がないため、より情報量が多い初期データを収集することができる可能性が高い。また、D最適計画は収集データ数も自由に設定できるという利点もあり、時間やコ

ストの制限を受けることも多い実験には適している。

2.2 回帰モデル, 分類モデルの作製

すでにデータ数が多い場合は回帰モデルや分類モデルを作製し, そのモデルを逆解析することで, 最適化することが可能となる。ここでいう逆解析とは探索空間内で順解析を繰り返すことで, 最適な条件を見つけるという逆解析である。逆解析による最適化の場合は探索条件がモデルの適応範囲内なのかを考慮することが重要である。また, モデルの作製にはどのような特徴量を使用するか, なども重要となるため, 解析対象材料(製品)に応じたノウハウを蓄積していく必要がある。

2.3 モデルの解釈

一般的にモデルの予測精度と解釈性はトレードオフの関係となる。例えば, 深層学習などの予測精度が一般的に高いモデルはブラックボックスモデルとなってしまう。MIでは「予測」よりも「解釈」に重きが置かれることも多い。そこで, 予測精度はあまり期待できないが, 解釈性が高い決定木なども活用される。また, 最近ではLIME, SHAP, Partial dependence plot, Permutation importanceなどXAI(説明可能なAI, 解釈可能なAI)技術の活用も増えてきている⁷⁾。ここで重要なのは, 得られるものはあくまでモデルの解釈であり, 因果関係ではないということである。解釈した結果が正しいかどうかは追加実験などを実施し, 別途検証する必要がある。

2.4 パーシステントホモロジー

パーシステントホモロジーとは, トポロジーの概念を利用した数学的解析手法であり, 図形の孔や空洞, 連結部分など, 構造の「つながり方」を定量的に抽出することが可

能である。例えば, ガラスの原子配置におけるリング構造の階層的関係の解析などに用いられており, 従来は発見できなかった構造を明らかにすることができている^{8) 9)}。材料開発では様々な画像データや分子動力学などから得られた構造情報などを取り扱うことが多く, パーシステントホモロジーへの期待が高まっている。

3. シース材料の配合最適化

3.1 配合の実験計画立案

当社で製造しているあるケーブルのゴムシース材料で, 引張強度と伸びが規格に対し裕度がないという課題があった。また, 同時期にベースゴムの供給不安が重なったことで, 引張強度と伸びを改善する代替配合を早期に確立する必要が生じた。今回のシースは14種類の配合剤から構成されており, 例えばそれぞれを3水準で試したとしても組み合わせは約478万通り(3^{14})もあり, 膨大な実験範囲となってしまう。また, 製造に関わる特性や強度以外の特性など様々な規格を満たしながら, 引張強度と伸びを改善していく必要があることから, この配合開発は多目的最適化となる。これらの特性にはトレードオフの関係となるものもあり, 従来は担当者の経験と勘から実験計画を考え, 多数の実験を行う必要があった。そこで, 実験計画立案にベイズ最適化を活用することとした。

図2はある時点における実験データを解析した際の獲得関数(次に実験する条件の価値)を示している。UMAPという手法を用いて, 探索空間である多次元の実験条件を2次元に次元削減し, 可視化したものである。丸印の位置は実験条件を, 大きさと色は獲得関数の大きさを示している。星印は実験済みデータの実験条件を示している。探索空間に対する実験済み条件の分布を確認することが可能で

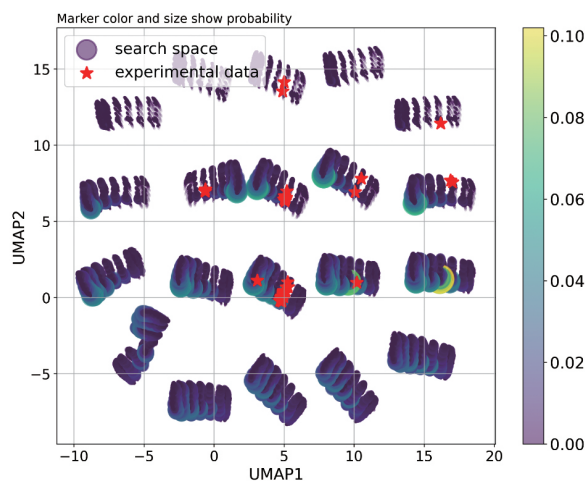


図2 探索空間内の獲得関数の可視化

Fig. 2. Visualization of the acquisition functions in the search space.

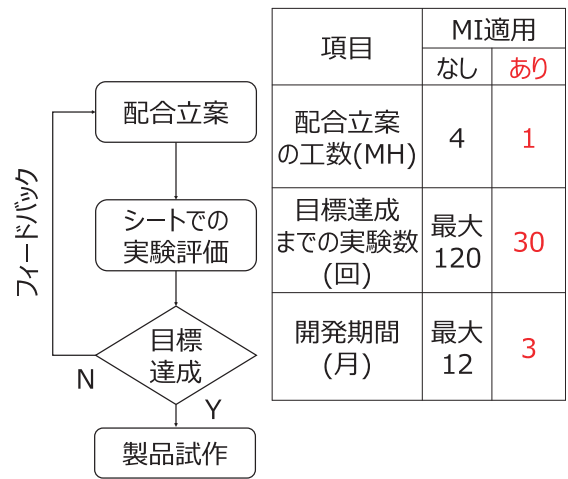


図3 配合開発のMI適用比較

Fig. 3. Comparison of Compound Development with and without MI Application.

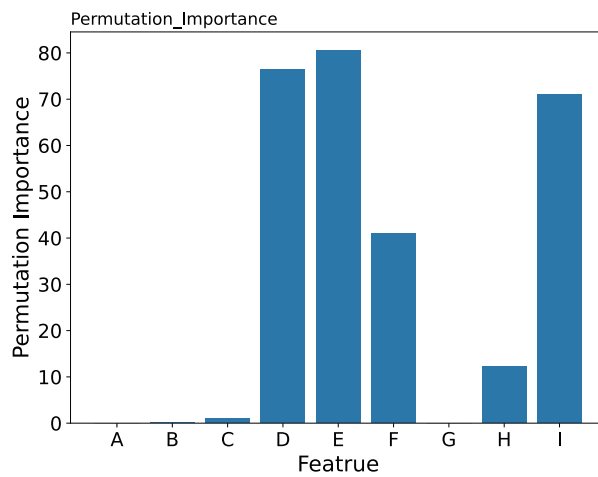


図4 配合材の重要因子特定

Fig. 4. Identification of critical factors for compounding materials.

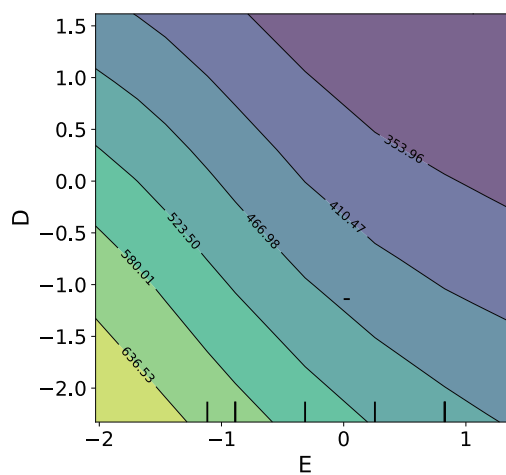


図5 モデルが考える配合剤D, Eに対する伸びの期待値

Fig. 5. Correlations between D and E compounding agents and elongation considered by the model.

ある。この中で獲得関数が最も高い条件を次の実験条件とする。その結果を用いて再度モデリングし、獲得関数を更新する。これを繰り返すことで最適化していく。その結果、約30回の実験回数で最適化を実施することに成功した。図3は従来手法とMIを活用した手法を比較した結果を示している。MIを活用することで開発期間を大幅に短縮できたことがわかる。また、今回のベイズ最適化過程において、担当者が考えていなかった条件が提示されるもあり、実際に配合してみたところ、比較的良好な特性を得られることが判明した。このようにMIは人に新たな気づきや考えるきっかけを与えるツールとしても活用できることを改めて認識した。

3.2 モデルの解釈

上記の配合開発では、ベイズ最適化に使用した確率モデルの解釈についても並行して実施した。図4に伸びを目的変数とした確率モデルをPermutation importanceで解釈した結果を示す。この図は各配合剤がモデルの予測値に与える影響の大きさを示しており、配合剤D、E、Iの重要度が高いことがわかる。そこで、配合剤D、Eが伸びに対してどのような関係となるのかをPartial dependence plotにて確認した。結果を図5に示す。周辺化により対象としている特徴量以外の影響を除去することで、対象としている特徴量と目的変数の相関関係を求めている。配合剤D、Eに対し、伸びは負の相関があることがわかる。D、Eが重要であることや負の相関があることはドメイン知識と一致しており、MIの有効性を裏付ける結果を得ることができた。

4. むすび

シーラ材料の開発にMI手法を活用した。実験条件立案にベイズ最適化を用いることで、開発期間を大幅に短縮することに成功した。解析結果は次元削減手法であるUMAPを用いることにより、可視化した。さらに、Permutation importanceやPartial dependence plotを用いることでモデルの解釈も実施した。今後もMIを活用し、材料技術開発の高効率化に貢献する。

参考文献

- 1) Y. Mo, et al., "First Principles Study of the Li10GeP2S12 Lithium Super Ionic Conductor Material", Chemistry of Materials 24 (1), pp.15-17, 2011
- 2) 独立行政法人 科学技術振興機構研究開発戦略センター, CRDS-FY2013-SP-01, データ科学との連携・融合による新世代物質・材料設計研究の促進 (マテリアルズ・インフォマティクス)
- 3) 中山ほか:「マテリアルズ・インフォマティクスによる全固体電池材料の最適化」, 表面技術, 第2号, pp.84-90, 2021
- 4) 松井ほか:まてりあ, 日本金属学会, PP.12-16, 2019
- 5) S. Pruksawan, et al., "Prediction and optimization of epoxy adhesive strength from a small dataset through active learning", Science and Technology of Advanced Materials, VOL. 20, 1010-1021, 2019
- 6) T. Morishita, et al., "Initial Sample Selection in Bayesian Optimization for Combinatorial Optimization of Chemical Compounds", ACS Omega, 8, pp.2001-2009, 2023
- 7) 亀谷:「説明可能AI技術のこれまでとこれから」, 電子情報通信学会 基礎・境界ソサイエティ Fundamentals Review, 16巻, 2号, 2022
- 8) Y. Hiraoka, et al., "Hierarchical structures of amorphous solids characterized by persistent homology", Proceedings of the National Academy of Sciences, 2016
- 9) Y. Onodera, et al., "Structure and properties of densified silica glass: Characterizing the order within disorder", NPG Asia Materials, 2020