

マテリアルズ・インフォマティクスを用いた 高分子複合材料の弾性率の予測モデル構築

Materials Informatics Approach to Predictive Models for Elastic Modulus of Polymer Composites

池田 祐子*
Yuko IKEDA

奥山 倫弘*
Michihiro OKUYAMA

中澤 幸仁*
Yukihito NAKAZAWA

押山 智寛*
Tomohiro OSHIYAMA

要旨

高分子複合材料の開発では、所望の弾性率を実現させる高分子樹脂、フィラー及び添加剤の組み合わせの探索範囲は膨大であるため、効率的に材料を選択できる開発プロセスが求められる。我々は、弾性率の予測モデルの構築に、材料開発プロセスの革新技術として注目を集めているマテリアルズ・インフォマティクス（以降MIと称す）を適用し、その適用範囲を考察した結果について報告する。

MIの適用では、高分子複合材料を構成するポリプロピレン（PP）、フィラー及び添加剤の各銘柄を0と1で組み合わせた記述子、フィラー及び添加剤は含有率を記述子とすることで説明変数を設定した。この手法は、各材料の物性値を用いる必要がないため、全ての実験処方に対して欠損なく説明変数を揃えることができる。このように設定した説明変数を用いて、PLS回帰により弾性率の予測モデルを構築した。構築したモデルの適用範囲を検証するために、新たに選択した9水準に対して弾性率を測定して実測値と予測値を比較した結果、概ね1,700MPaから3,100MPaの範囲で残差を300MPa以内に収めることができた。また、フィラーの含有率が高い領域で予測精度が低い理由を考察した。

本稿で示したMIに基づく予測モデルは、膨大な材料候補の中から迅速に所望の物性を有する材料選択が必要となる高分子複合材料の開発において有用と考えられる。今後は、高弾性率として予測される未実験の処方が提案できるように、実験データを増やす中で予測モデルを改良し、高弾性率側、高フィラー含有率側の予測精度の向上を目指していく。

Abstract

In development of polymer composites, there is a need for an advanced development process since there are an enormous number of possible combinations of fillers and additives to be searched to realize a desired polymer. Recently, materials informatics (MI) has been focused on the data-driven approach to find novel materials or a suitable combination of materials from material data sheets.

We carried out a materials informatics approach on predictive models for elastic modulus of polymer composites and have constructed a predictive model. We have also specified the applicability domain for a rapid selection of materials with desirable elastic modulus.

In application of MI to existing experimental data, we described the explanatory variables by a combination of 0 and 1 representing polymer, filler, and additive or by the content rate of filler and additive, without using the property data of materials. To validate the predictive model, comparisons were made between measured elastic moduli and predicted ones for nine levels of polymer composites, and it was found that the residual error was less than 300 MPa for a range of 1,700 MPa to 3,100 MPa. We also considered why the accuracy of prediction is low in the range of high content rate of filler.

We have constructed a predictive model for the elastic modulus of polymer composites by using a partial least square (PLS) regression model. The MI-based predictive model of this paper is useful for discovering a suitable combination of polymer, filler, and additive with desirable elastic modulus. We have a plan to carry out further experiments and increase the number of data points, especially in the range of high elastic modulus or high filler content, to establish a more accurate predictive model.

* 開発統括本部 要素技術開発センター 機能材料開発室

1 序論

高分子樹脂にフィラー及び添加剤を均一に分散させることで所望の機械特性を発現できる高分子複合材料は、軽量かつ高強度な性能が着目され、航空機や自動車産業において金属に代わる次世代の材料として実用化が進んでいる。所望の機械特性を実現させるには、材料種の選択と組合せ、混合比及び複合化のプロセス条件等を適切に制御し最適化する必要がある。しかしながら、選択する材料種とその組合せの数は膨大であるため、その全てを実験することは不可能である。また、技術者の勘と経験に基づく処方設計のみでは多くの開発費用と時間を要する。

近年、材料開発プロセスの革新技術として注目を集めているMIは、データ駆動により新材料を創出する手法であり、無機系の固体電解質材料等の発見の他¹⁻³⁾、低分子の有機発光材料の探索⁴⁾や溶解度予測⁵⁾に適用され注目を集めている。一方、高分子材料は、MIの適用が遅れている。その理由は、分子構造や物性に関して、同条件の実験データを揃えることが困難なこと、温度や混練条件等のプロセス条件による影響が大きいことが挙げられるが、以下に述べる先駆的な適用事例もある。

物質・材料研究機構 (NIMS) の高分子データベース (PoLyInfo)⁶⁾からガラス転移温度等の実験データを取得して、分子構造の情報を数値化する手法であるフィンガープリントを用いて得られる記述子を説明変数とし、機械学習を駆使して目的の物性値を予測することで、新しい高分子材料の提案が行なわれている⁷⁾。一方、ポリプロピレン (PP) とタルクからなる高分子複合系の文献データを利用した例では、PPの平均分子量、タルクの混合比等の物性データを説明変数とし、高弾性率に寄与する物性を特定している⁸⁾。これらの事例では、解析するテーマと取り扱うデータ数に応じてMIを的確に利用することで価値の高い成果につなげている。本稿では、上述した先駆的な研究例とは異なる観点から、高分子複合材料の開発にMIを適用した。

本稿の目的は、高分子樹脂としてPPを用いた複合材料にMIを適用して、迅速に高弾性率につながる最適処方を提案することである。そのためには以下に述べる二つの手順が必要である。第一の手順は、既存の実験データを用いて弾性率の予測モデルを構築することである。第二の手順は、材料種及び含有率から考えられる全ての複合材料の組み合わせを挙げて、第一の手順で構築した予測モデルを用いて弾性率を予測し、高弾性率の処方を提案し、実験で検証することである。この二つの手順を繰り返すことで、MIから導いた提案の確度を上げることができる。

第一の手順を構築する上での課題は、同一のモノマーからなる多種多様な銘柄が存在するという高分子材料の特性が挙げられる。例えば、PPには分子量分布、共重合度等の数値データに加えて、立体規則性 (タクティシティー) を考慮する必要がある。そのため、分子構造を

フィンガープリントで表現する手法を用いることが困難であると共に、物性データを蓄積するには膨大な工数を要し、温度や混練条件等のプロセス条件による物性への影響も大きい。この特性は、フィラーや添加剤も同様である。そこで我々は、上記の課題を解決して、保有している全ての実験データを回帰モデル構築に反映させるために、各材料種の銘柄と含有率を用いた記述子、つまりダミー変数を説明変数として用いることを考えた。本稿の第2節では、各材料種の組み合わせと含有率から記述子を設定する手段と、それを説明変数に用いた弾性率の予測モデルの構築手順及び結果について述べる。第3節では、構築した予測モデルの適用範囲について考察する。最後に、高分子複合系にMIを適用した解析結果のまとめと、今後の展望を述べて結びとする。

2 弾性率の予測モデルの構築

2.1 各材料種の銘柄と含有率を用いた記述子の設定

本稿では、PP、フィラー及び添加剤としてそれぞれ11種類、18種類、20種類の材料の組み合わせと含有率からなる180種類の弾性率の実験データを用いた。説明変数に用いるPP、フィラー及び添加剤の分子構造や熱物性値等の特性値は、データが欠損しているものがあり、その課題を克服する手段として、特性値に代わる新たな説明変数の設定を試みた。

我々が設定した説明変数は、実験処方におけるPP、フィラー及び添加剤の使用の有無を0と1のダミー変数で記述すると共に、含有率と合わせてベクトルで表現した。これにより、全データを解析に使用することができるようになった。

まず、PPの説明変数は、以下の式(1)によって記述した。ここで、PPの銘柄は p_i ($i=1$ から11)と定義した。

$$x_{p_i} = \begin{cases} 1 & (i=\alpha) \\ 0 & (i \neq \alpha) \end{cases} \quad (1)$$

また、フィラー及び添加剤の説明変数は、以下の式(2)及び式(3)によって記述した。ここで、フィラー及び添加剤の銘柄は、それぞれ f_i ($i=1$ から18)、 a_i ($i=1$ から20)と定義した。 c_{fi} 、 c_{ai} はそれぞれの含有率である。

$$x_{f_i} = \begin{cases} c_{f_i} & (i=\alpha) \\ 0 & (i \neq \alpha) \end{cases} \quad (2)$$

$$x_{a_i} = \begin{cases} c_{a_i} & (i=\alpha) \\ 0 & (i \neq \alpha) \end{cases} \quad (3)$$

上記式(1)から(3)を用いて、例えば下記式(4)のようなベクトル表現で説明変数を設定した。下記式(4)の場合であれば、 p_1 のPP、 f_1 のフィラーを10%、 a_{20} の添加剤を5%含有させた高分子複合材料であることを示す。

$$x = \begin{bmatrix} x_{p_1} \\ \vdots \\ x_{p_{11}} \\ x_{f_1} \\ \vdots \\ x_{f_{18}} \\ x_{a_1} \\ \vdots \\ x_{a_{20}} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 0 \\ 10 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 5 \end{bmatrix} \quad (4)$$

2.2 弾性率の予測モデルの構築手順及び結果

高分子樹脂、フィラー及び添加剤から構成される高分子複合材料について、式(1)から式(3)で定義した計49個の説明変数を用いて設定することができたので、弾性率の予測モデルは部分的最小二乗回帰 (Partial Least Squares regression, PLS回帰)⁹⁾を用いて構築した。PLS回帰は、Python3の機械学習ライブラリである scikit-learnを用いて行った。PLS回帰では、説明変数間が互いに無相関となるよう線形に変換した潜在変数を用いる。

まず保有しているデータを用いて、予測モデルの精度を評価した。180個のデータを学習データが85%、試験データが15%程度の割合になるよう、ランダムにそれぞれ153個と27個に分割したデータセットを作成した。学習データを基にPLS回帰で予測モデルを構築し、その予測モデルを使用して試験データの予測値を算出した。これは予測値と実測値を比較し、未知のデータ(試験データ)に対して予測モデルの適応度を評価する手法である。潜在変数の数は、以下の式(5)を用いて表される試験データにおけるRoot Mean Square Error (RMSE, ε_1)が最小となる値を用いた。

$$\varepsilon_1 = \sqrt{\frac{\sum_{i=1}^N (y_{obs,i} - y_{pred,i})^2}{N}} \quad (5)$$

ここで式(5)において、 $y_{obs,i}$ は試験データ*i*における実測値、 $y_{pred,i}$ は学習データから構築した予測モデルを適応して試験データ*i*について算出した予測値である。

上記式(5)を用いて ε_1 を算出した結果、今回は潜在変数が13個の場合を採用した。その結果、試験データにおける ε_1 は156MPa、決定係数 r^2 は0.95となった。以上の結果から、構築した予測モデルは概ね良好であると言える。ただし、潜在変数の数が13個と多いことが懸念事項として挙げられる。また、これらの結果はランダムに分割した1つのデータセットの試験データのみに対する精度を示しており、汎化性の有無を判断することはできない。

そこで次に、上記の課題を解消させるために、Leave-One-Out交差検証(LOOCV)を行った。LOOCVでは、ある1つのデータを除いた179個を学習データとし、PLS回帰を用いて予測モデルを構築した。構築した予測モデルを用いて残り1つの試験データの予測値を算出し、予測値と実測値の誤差を算出する。これを180回繰り返すこ

とで、所持しているデータ内での汎化性を確認することが可能である。LOOCVに関しても、潜在変数の数は以下の式(6)を用いて求まる ε_2 が最小となる値を用いた。

$$\varepsilon_2 = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (6)$$

ここで式(6)において、 y_i はデータ*i*における弾性率の実験値、 \hat{y}_i はデータ*i*を除いた179個のデータで学習し、除いたデータ*i*でテストして算出した弾性率の予測値である。

Fig. 1に潜在変数の数と ε_2 の関係を示す。Fig. 1から、潜在変数の数が4個の場合で ε_2 が最小値(309MPa)を示した。その時の、決定係数 r^2 は0.73であった。以降は、LOOCVによって構築した予測モデルを基に議論を進める。

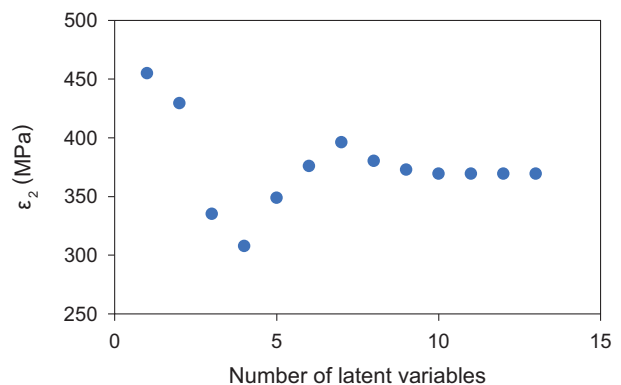


Fig. 1 Relationship between predicted ε_2 and the number of latent variables.

The minimum ε_2 is 309 MPa when the number of latent variables is 4.

次に、4個の潜在変数を用いて弾性率の予測モデルをPLS回帰で構築した。LOOCVによって構築した予測モデルから算出した実験値と予測値の関係をFig. 2に示す。Fig. 2から、構築したモデルは、定性的にみれば、高弾性率側で残差が大きくなる傾向を示しているが、概ね良好なモデルと言える。

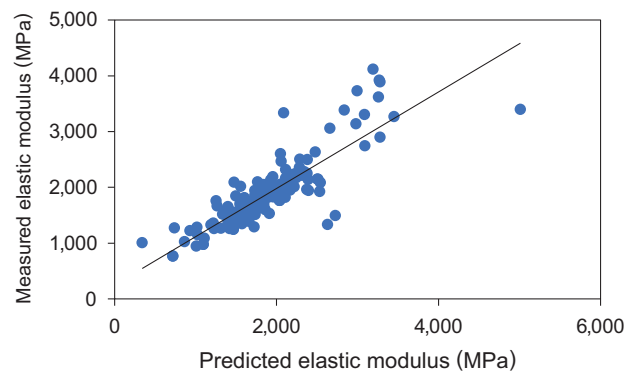


Fig. 2 Relationship between measured elastic modulus and modulus predicted by LOOCV (Leave One Out Cross Validation) using four latent variables.

Although the model constructed by the LOOCV tends to show a large residual error on the high-modulus side, the model is appropriate.

今回の解析では、説明変数として、連続値ではなくPP、フィラー及び添加剤の各銘柄の使用の有無を基に49個のダミー変数で記述している。このような説明変数を用いても、Fig. 2 に示されるように良好なモデルを作成することができた。ただし、高弾性率側でのデータ密度が相対的に低いので、モデルの適用範囲を明確にする必要がある。

3 弾性率の予測モデルの考察

PLS回帰による予測モデルを適切に運用するために、その適用範囲について、以下の2点を明らかにするための考察を行った。すなわち、(Q1)モデルの構築に用いたデータから残差を可視化することで、予測精度を定量化しモデルの適用範囲を明らかにしていくこと、(Q2)予測精度の低い領域は、残差の大きくなる理由について考察し、モデルの改良に活かしていくこと、である。

3.1 構築した弾性率の予測モデルの適用範囲の特定

3.1.1 構築した予測モデルの残差からの適用範囲の特定

予測モデルの適用範囲を考察するためには、各弾性率における実験値と予測値の誤差を算出する必要がある(Q1)。そこで、前節で示したFig. 2から各弾性率における実験値と予測値の残差の絶対値を算出し可視化した。その結果をFig. 3に示す。

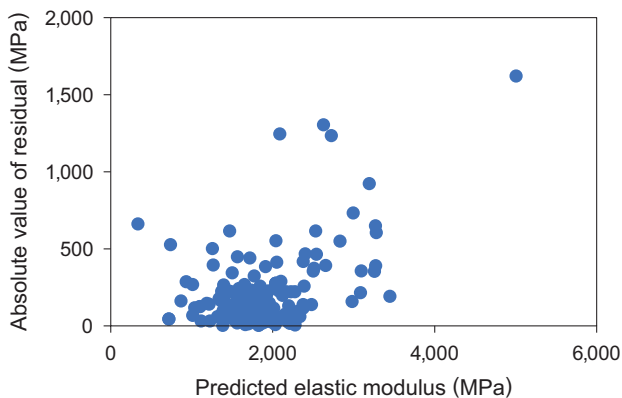


Fig. 3 Relationship between absolute value of residual error and predicted elastic modulus.
The percentage of the data having residual error of 300 MPa or less is 90% in the range of 2,500 MPa or less and 15% in the range of 2,500 MPa or more.

第2節(2)のLOOCVを用いて算出された ε_2 が309 MPaであったことを考慮し、残差が300 MPa以下のデータが存在する範囲と割合の偏りを求めた。その結果、2,500 MPa以下の範囲で残差を300 MPa以内に収めることができていた割合が90%、2,500 MPa以上では、15%であった。高弾性率側で予測精度が低くなっているが、これはデータ密度が小さいことが要因であるため、実験データを増やして予測精度を上げていく必要がある。

3.1.2 追加実験からの適用範囲の特定

次に、予測した一部の高分子複合材料について、実測を行い予測値と比較する側面からも、構築したPLS回帰の精度の検証を行った。今回MIに適用した180種類以外にも、PP、フィラー及び添加剤の組み合わせは数多く考えられる。そこで、考え得る57万通りの高分子複合材料について、構築した回帰式を用いて弾性率の予測値を算出した。この予測値を基に、材料の偏りや予測値の領域に偏りが出ないように9水準を抜粋し、実測を行った。実測候補の水準を抜粋する際は、以下の2点を配慮して選択をした。第2節で予測モデル構築に用いた、保有している実測データを基に49変数の x 次元が(i)近い、(ii)遠いと考えられる水準が混在するように選択した。(i)からは、保有データと近いと考えられる領域での未知データに対して、構築した予測モデルの予測精度は正しいか、(ii)からは、保有データと遠いと考えられる領域での未知データの予測は可能であるかを検証することが目的である。

実測値と予測値を比較した結果をFig. 4に示す。Fig. 4に示すように、予測値が1,684~3,148 MPaの範囲にある水準B~H ((i)に分類)の残差は300 MPa以内に収めることができた。これは第2節(2)で述べた、LOOCVでの ε_2 (309 MPa)と非常に近く、構築した予測モデルの精度が妥当であることを示している。一方で、我々が予測したいと考える高弾性率(水準I, (ii)に分類)の残差は3,017 MPaと大きく、予測精度が低いことが分かる。つまり、保有データと x 次元が遠いと考えられる領域については、予測が困難である可能性を示唆している。また、極端な低値として予測された水準A ((i)に分類)についても、残差が762 MPaと予測精度が低い。これは、Fig. 2からも分かるように、低弾性率領域のデータ密度が低いことが原因であると考えられる。

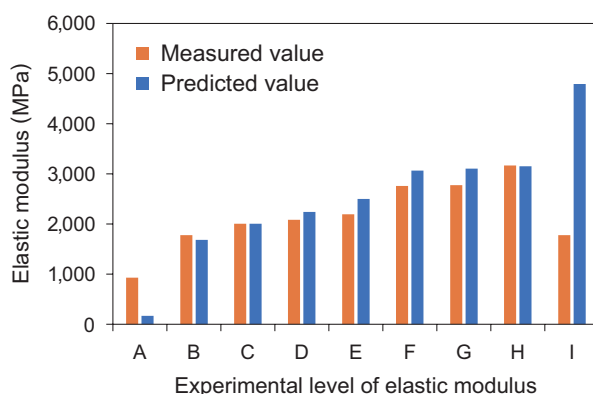


Fig. 4 Comparison of elastic modulus between measured value and predicted value for nine levels.

The residual error is 300 MPa or less for levels B to H (predicted value is 1,684 to 3,148 MPa) but is over 3,000 MPa for level I.

以上、構築した弾性率の予測モデルと、追加の検証実験の二つの側面の検証からも分かるように、保有データと x 次元が遠い、または3,000 MPa以上の高弾性率領域の予測精度が低いことが分かる。精度向上のためのモデ

ル改良に活かすために、高弾性率領域の予測精度が低くなる原因を明確にすることを考えた。

3.2 高弾性率領域における予測精度低下の原因の究明

高弾性率領域の予測精度が低くなった原因を解明するため (Q2), 保有データの内, 1つのデータ点に着目して解析をした。Fig. 2 において, 予測値が高弾性率領域に属し, 実測値との残差が1,609MPaと最も精度が低く予測されたデータ (以降 δ と称す) が, 上記の特性を捉えたデータである。すなわち, δ の予測精度が低下する原因を解明することでモデルの改良に繋がると考えた。

まず, δ の説明変数を確認してみるとフィラー f_4 の含有率が40%と, その他の高分子複合材料と比較して, フィラー含有率の高い高分子複合材料であることが分かった。一般に, フィラーの含有率を高めると弾性率も高まるが, 適正点が存在することが知られている。今回, フィラー含有率が高い高分子複合材料であるデータ δ の残差が大きいため, この適正点が正確に予測できていない可能性があると考えた。そこで, フィラー f_4 を含有しているデータを抽出し, 含有率と弾性率の実測値と予測値の関係を可視化した。

180個のデータから, フィラー f_4 の含有率が, 40% (δ), 30%, 20%, 10%, 1%のデータを抽出した。同含有率の中に, 異なるPPを使用したデータが複数個存在する場合は, 中央値を取って評価した。以下のFig. 5に弾性率の実測値と予測値, それぞれの中央値とフィラー f_4 の含有率の関係を示す。

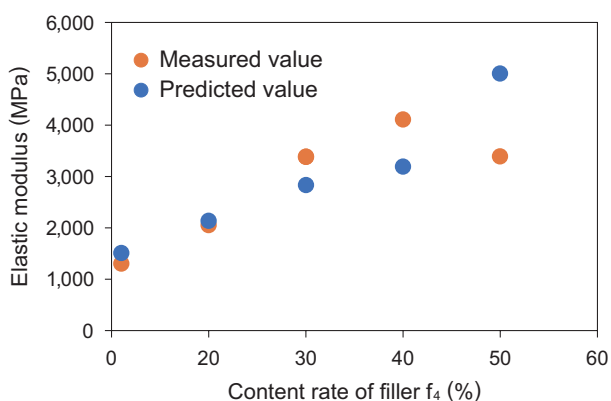


Fig. 5 Relationship between elastic modulus (measured value and predicted value) and content rate of filler f_4 .

The predicted elastic modulus has a linear relationship with the content rate of filler f_4 . In contrast, the measured elastic modulus deviate from the linear relationship.

Fig. 5 から明らかなように, 弾性率の予測値とフィラー f_4 の含有率の間には, 直線的に増加していく傾向がみられる。対して実測値では, その直線から外れる挙動を示し, フィラー f_4 の含有率が40%以上では, 弾性率の予測値が実測値を下回っていることが分かる。

以上の考察から, フィラー f_4 の含有率が40%以上では, 弾性率の予測精度が低下する。Fig. 4 で示した, 水

準Iにおいてもフィラーの含有率は60%と, フィラー含有率の高い高分子複合材料であった。水準Iで使用したフィラーは f_{10} であることから, フィラーの種類に関わらず高弾性率側での予測精度は低下する可能性がある。最適なフィラーの含有率を予測するためにも, 高弾性率側の予測精度の改良が必要である。今後, フィラー含有率が高い領域の予測精度を高めるためにも, 説明変数に非線形項を導入する手段や, 非線形のモデルを活用する等により, 改善を図っていく。

4 結び

本稿では, 11種類のPP, 18種類のフィラー及び20種類の添加剤の180個の組み合わせからなる実験データに対してMIを適用し, 弾性率の予測モデルを構築し, その妥当性を検証した。その結果, 高分子複合材料の弾性率に関して, 既存の実験データの銘柄と含有率を説明変数として用いた予測モデルの構築方法とその結果及び適用範囲を示すことができた。上記の手法は, MIを進める上で課題となる数値データの欠損を克服する手段の一つとなりうることを確認できた。データ全体でみれば, データ密度が相対的に低い領域に課題はあるが, 比較的良好なモデルを作成することができ, 所望の弾性率を予測する有用なモデルが構築できた。

構築したモデルの適用範囲を検証するために, 新たに選択した9水準に対して弾性率を測定して実測値と予測値を比較した。その結果, 概ね1,700MPaから3,100MPaの範囲で残差を300MPa以内に収めることができた。この残差は, LOOCVでの ε_2 (309MPa) と非常に近く, 構築した予測モデルの精度が妥当であることを示した。また, 予測精度が低い, フィラーの含有率が高い領域では実験データを増やすと共に, 説明変数に非線形項を導入する手段等により改善を図っていく。

今後は, 高分子複合材料へのMI適用による, 迅速な最適処方提案のために, 高弾性率として予測される未実験の処方を提案し, 実験データを増やす中でモデルを改良していく。このように実験と解析を繰り返しながら, 高弾性率側, 高フィラー含有率側の予測精度の課題を改善する中で, 処方選択幅を広げていく。本稿で示した, 既存の実験データの銘柄と含有率を説明変数として用いた予測モデルの構築手法は, データ不足や欠損に課題を抱える他のテーマにも応用可能であり, 今後広く活用していく。

謝辞

本稿をまとめるにあたり, 数多くの議論をしていただき有用なご指摘をいただいた東京大学の船津公人教授, 及びジョージア工科大学のMartha Grover 教授に感謝します。

●参考文献

- 1) R. Jarem, K. Kanamori, I. Takeuchi, M. Nakayama, H. Yamasaki and T. Saito, *Sci. Rep.*, 8, 5845, (2018).
- 2) F. Ren, L. Ward, T. Williams, K. J. Laws, C. Wolverton, J. Hattrick-Simpers and A. Mehta, *Sci. Adv.*, 4, eaaq1566, (2018).
- 3) R. Yuan, Z. Liu, P. V. Balachandran, D. Xue, Y. Zhou, X. Ding, J. Sun, D. Xue and T. Lookman, *Adv. Mater.*, 30, 1702884, (2018).
- 4) R. Gomez-Bombarelli, J. Aguilera-Iparraguirre, T. D. Hirzel, D. Duvenaud, D. Maclaurin, M. A. Blood-Forsythe, H. S. Chae, M. Einzinger, D. Ha, T. Wu, G. Markopoulos, S. Jeon, H. Kang, H. Miyazaki, M. Numata, S. Kim, W. Huang, S. Hong, M. Baldo, R. P. Adams and A. AspuruGuzik, *Nat. Mater.*, 15, 1120, (2016).
- 5) 鈴木天音, 木倉悠一郎, 田中健一, 船津公人, *Journal of Computer Chemistry, Japan*, 19, 1, (2018).
- 6) PoLyInfo: <http://polymer.nims.go.jp>
- 7) 山田寛尚, W. Stephen, C. Liu, 吉田亮, 統計関連学会連合大会 (2018).
- 8) M. McBride, N. Persson, E. Reichmanis and M. A. Grover, *Processes*, 6, 79, (2018).
- 9) S. Wold, M. Sjöström and L. Eriksson, *Chemom. Inetll. Lab. Syst.*, 58, 109, (2001).