

機械学習を活用した放射線グラフト重合率の予測

国立研究開発法人 量子科学技術研究開発機構
植木 悠二*, 瀬古 典明, 前川 康成

Conventional polymer material development relies on inefficient trial-and-error experiments based on the experience and intuition of researchers. Therefore, the development of new polymer materials requires an enormous amount of time and high costs. Furthermore, conventional polymer material development has problems such as the difficulty of responding immediately to increasingly sophisticated and diverse social demands. As a solution to these problems, the use of materials informatics (MI), which incorporates data science such as artificial intelligence (AI) and machine learning into the traditional empirical experimental science, has recently become a promising approach. This article reports on the development of an AI model that can instantly predict the grafting yields based solely on the chemical and physical features of the monomers used in the radiation-induced graft polymerization, as an example of polymer material development utilizing machine learning.

Keywords: radiation-induced graft polymerization, machine learning, grafting yield, prediction

1 はじめに

安価で加工しやすいプラスチックのような高分子材料は、私たちの暮らしに必要な不可欠な材料であり、新しい機能を持つ高分子材料の発見・開発が社会や産業の発展に大きく寄与してきた。高分子材料の機能は、原料となる薬品（モノマー）の種類や組成、反応条件などの様々なパラメーター（因子）が複雑に絡み合う

ことにより発現するが、それらパラメーターの組み合わせは無限に存在する。そのため従来型の高分子材料開発では、研究者や熟練技術者の「経験や勘」、あるいは「偶然」を頼りに、非常に多くのパラメーターを組み合わせた地道な検討が繰り返し行われてきた。しかし、このような非効率的な試行錯誤をともなう材料開発では、非常に長い開発期間と多大な研究費が必要となる。また、求められる機能や性能が高度化・多様化する現在において、所望の機能を発現させる最適解（最適組成比や最適実験条件など）を経験や勘だけを見つけて出すことは、ますます困難な状況となつつある。さらに、近年では製品のライフサイクルは短縮化する傾向にあり、これまでの「実験至上主義」的な方法論では材料開発スピードが追いつかないことも事実である。それゆえ、所望する機能・性能を有する高分子材料を正確かつ迅速に開発する新しい技術や方法論の確立が切望されている。

上記の問題を解決する手段として、近年、従来の経験的な実験科学に人工知能（artificial intelligence, AI）や機械学習（machine learning）などのデータ科学を取り入れた「マテリアルズ・インフォマティクス（materials informatics, MI）」が注目されている。AIは入力されたデータに潜む法則性（ルール）やパターンを高速に見つけ出すことができるため、実験結果の事前予測や最適解の提案が可能となる。そのためAIの活用は、試行錯誤的な実験工程を不要とする低コストで迅速性のある効率的な材料開発が可能となるだけでなく、これまで研究者が見逃していた有益な科学的知見の発見や新規材料の創製に繋がるものとして期待されている¹⁾。

これまでもMIの活用により、リチウム超イオン伝導体²⁾、超長寿命リチウムイオン電池³⁾、超低熱伝導物質⁴⁾、金属ガラス⁵⁾、熱電変換材料⁶⁾など幾つかの革新的材料や、新規合成経路⁷⁾が発見されている。また、MIは、リチウムイオン電池の寿命⁸⁾、触媒反応の収率⁹⁾、NMRの化学シフト値¹⁰⁾など、様々な予測にも利用されている。

一方で、画像認識、機械翻訳、囲碁・将棋AIなどの

Prediction of radiation-induced graft polymerization yield using machine learning.

Yuji UEKI*, Noriaki SEKO and Yasunari MAEKAWA (National Institutes for Quantum Science and Technology),

〒370-1292 群馬県高崎市綿貫町 1233

TEL: 027-346-9202, E-mail: ueki.yuji@qst.go.jp

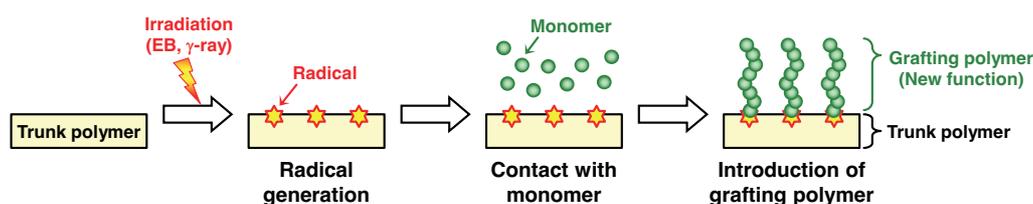


Figure 1. Schematic diagram of radiation-induced graft polymerization.

ビッグデータを活用する機械学習のひとつである深層学習 (deep learning) は、すでに人間を凌駕した存在になっており、社会に普及している。そのため、材料開発分野においても、AI にデータを入力すれば、自動的に所望の機能を有した新素材を創製してくれる「魔法の杖」が存在しているかのようなイメージがあるが、実際はそのような万能 AI はまだ存在していない。特に高分子材料開発では、

- 材料データの収集に多大なコストがかかる。
- 予測したい性能や特性が多岐にわたる。
- 画一的な実験手法が確立していない。

などの理由から機械学習に使用できるデータ数はたったの 10 件–100 件程度という場合が多く、現状の高分子材料開発分野の MI では、「極めて少ないデータ数をもとに、如何にして高精度に材料物性を予測して、所望の機能を有した新高分子材料を開発するか」というパラドキシカルな挑戦となっている。そのため、予測したい目的に無関係なデータだけを AI に学習させても、作成された AI モデルは誤った予測をするだけである。

現状、完全に「経験や勘」を排除したデータを用いて AI モデルを作成することは不可能であり、研究者の専門知識をもとに、あらかじめ山を張って (適切な事前知識を与えて) データを準備しなければならない。つまり、AI や機械学習は、あくまでも研究者をアシストする補助ツールであり、高分子材料分野の MI においては、専門知識を持った研究者が、どのようなデータを用いて、どのような結果を得たいか、その結果をどのように評価するか、何をもってゴールとするか等、課題設定や事前のルール設定を的確に行いながら材料開発を進める必要がある。

筆者らは、効率的な高分子材料の加工技術のひとつである放射線グラフト重合を活用した材料開発を進めている^{11,12)}。放射線グラフト重合の概略を Fig. 1 に示

す。この放射線グラフト重合は、ポリエチレンなどの汎用高分子材料 (基材) に電子線やガンマ線などの放射線を照射した後、その基材をモノマー溶液に浸すだけで、モノマーの機能が付与された新しい機能性高分子材料が創製可能な技術である。この技術では、使用するモノマーの種類により簡単に所望の機能を付与することができる。さらに、基材に既製の汎用品を使用することができるため、材料開発を一からする手間がなく、材料開発期間の短縮化が可能となる。

これらの利点を生かし、筆者らは近年、比較的短い期間において超強塩基性領域において機能する金属吸着材¹³⁾、環境水に含まれる放射性セシウム除去材¹⁴⁾、燃料電池自動車用高分子電解質膜¹⁵⁾などの開発に成功してきた。しかし、社会環境が大きく変化する現状において、材料開発競争は激化の一途を辿っており、今後も高い産業競争力を維持・強化していくためには、放射線グラフト重合研究においても AI や MI の利活用は必須な状況となりつつある。

本稿では、放射線グラフト重合研究における機械学習の有効性検証のファースト・ステップとして、グラフト重合反応に使用するモノマーの物性情報のみからグラフト重合率を瞬時に予測できる AI モデルの作成を試みたのでその結果について報告する¹⁶⁾。

2 実験

2.1 データ収集

2.1.1 放射線グラフト重合

機械学習とは、端的に言えば機械 (コンピューター) が既知データを学習することにより、その既知データに潜む規則性や判断基準などのパターンを表す関数モデルを導出し、その関数モデルを用いて未知データに対する予測を行う手法のことである。なお、機械学習では、予測したいデータ・値を「目的変数」、目的変

数を説明するために使用するデータ・値を「説明変数（または、記述子）」と呼ぶ。機械学習において、最も単純な説明変数が1種類の関数モデルは、 $y = ax + b$ の関数で表すことができ、このとき y 項は目的変数、 x 項は説明変数となる。つまり、既知データにより導出した関数モデル $y = ax + b$ は、説明変数 x によって特徴づけられており、この関数モデルの x 項に未知データの説明変数 x_i を入力すれば、未知データにおける目的変数 y_i を予測することができることとなる。また、説明変数が複数になれば関数モデルは多項式化し ($y = a_1x_1 + a_2x_2 + a_3x_3 + \dots + b$)、より複雑な数式となる。本研究では、予測したい目的変数をグラフト重合反応における反応率（グラフト重合率）と定義し、説明変数にはモノマーの物性情報として分子量や体積などいくつかの特徴量を用いた。

また、機械学習の方法としては、入力データ（説明変数）に対する正解情報を与えることで、未知試料の説明変数から目的変数を予測する「教師あり学習」と、正解情報を与えずに目的変数を予測する「教師なし学習」がある。本研究では、一般的に学習精度が高く、学習速度も早いといった利点を有する「教師あり学習」により、目的変数である「グラフト重合率」を予測する AI モデルの作成を試みた。

機械学習では、最初に機械に学習させるためのトレーニングデータ（または、学習データ、教師データと呼ぶ）と、テストデータを準備する必要がある。トレーニングデータとテストデータは共に、「入力データ（説明変数）」と「正解情報（目的変数）」をペアにしたデータ群のことであり、トレーニングデータは説明変数と目的変数との間に潜む法則性を表すモデル関数（AI モデル）を導出するために使用し、一方、テストデータは得られたモデル関数の妥当性を評価するために使用する。

本研究では、正解情報となるグラフト重合率（目的変数）は実際にグラフト重合実験により取得し、説明変数となるモノマーの物性情報は 2.1.2 目に示す量子化学計算より導出した。得られたトレーニングデータを機械に学習させることにより、グラフト重合率とモノマーの物性情報との間に潜む法則性を表すモデル関数（AI モデル）が自動的に作成される。このようにして作成した AI モデルに、グラフト重合率を予測したい未知モノマーの説明変数（テストデータの説明変数）を入力すると、そのグラフト重合率を予測することができる。未知モノマーの予測グラフト重合率が十分に優れていれば、AI モデルの作成は終了し、不十分であれば、AI を再度作成するために実験データを追加する

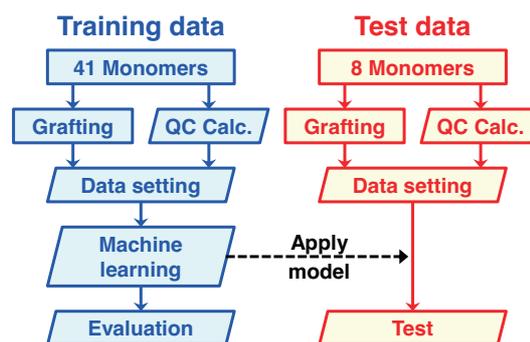


Figure 2. Simple analytical flow chart from data collection to the evaluation of predictive modeling and testing of the predictive model. (Reprinted some modifications from Ref. 16, with the permission of Elsevier publishing.)

などの処置をして予測性能の向上を図ることとなる。

本研究ではメタクリル酸エステル系モノマーを用いた電子線エマルジョングラフト重合反応をモデル反応として設定し、機械学習を利用して本重合反応の反応率（グラフト重合率）を予測する AI モデルの作成に取り組んだ。Figure 2 に、データ収集から予測モデルの評価、テストまでの簡単な分析フローチャートを示す。本研究では、分子構造の異なる 49 種類のメタクリル酸エステル系モノマーを用い、そのうち、41 種類をトレーニングデータ、残りの 8 種類をテストデータとした。目的変数であるグラフト重合率は、下記に示す電子線エマルジョングラフト重合反応実験により取得した。

基材となるポリエチレン製不織布（サイズ： $3 \times 3 \text{ cm}^2$ ）に、窒素雰囲気下において電子線を 50 kGy 照射した。その後、照射したポリエチレン製不織布を、各種メタクリル酸エステル系モノマーを含むエマルジョン水溶液（組成： 5 wt\% モノマー、 0.5 wt\% 界面活性剤（Tween 20）、 94.5 wt\% 水）に浸漬し、反応温度 $40 \text{ }^\circ\text{C}$ にて 1 時間加熱し、グラフト重合反応を実施した。本研究では、各種メタクリル酸エステル系モノマーにおいて、それぞれ 5 回の実験を実施した。目的変数であるグラフト重合率は、グラフト重合反応前後の基材の重量増加百分率から算出した。

2.1.2 量子化学計算によるモノマー物性情報の数値化

メタクリル酸エステル系モノマーを機械学習させるためには、モノマーの分子構造や物性情報を数値化し

たデータとして取り扱う必要がある。本研究では、説明変数として用いる各種メタクリル酸エステル系モノマーの物性情報は量子化学計算（計算ソフトウェア：Spartan'18, 計算手法：密度汎関数法, 基底関数： ω B97X-D/6-31G*）により数値化した。

本研究で使用した説明変数を Table 1 に示す。本研究では、モノマーの分子量や体積などの基本的な物性情報（11 種類：①–⑪）だけを説明変数として機械学習させた場合、予測精度の高い AI モデルを作成することができなかった。

そこで筆者らのこれまでの化学的知見に基づき、グラフト重合反応に寄与する可能性がある説明変数の選定を試みた。具体的には、基本的な物性情報（11 種類）に加え、各モノマーの共通部分構造であるメタクリル酸基を構成する 11 個の原子（炭素原子：4 個、水素原子：5 個、酸素原子：2 個）の電荷（11 種類：⑫–⑲）、各原子の NMR 化学シフト（11 種類：⑳–㉓）、および、メタクリル酸基の FT-IR スペクトルの波数とその強度（16 種類：㉔–㉙）の合計 49 種類の物性情報を説明変数として選定した。メタクリル酸基を構成する各原子のナンバリング、および、FT-IR スペクトルの種類は、Fig. 3 のように定義した。また、FT-IR スペクトルにおいて、“*v*”と“*i*”はそれぞれ波数と強度を意味し、たとえば C1=C2 二重結合における伸縮振動の波数と強度は、それぞれ「 $v_{C1=C2, str.}$ 」と「 $i_{C1=C2, str.}$ 」と表記した。

2.2 データセット構築と説明変数選択

機械学習に用いるデータセットは、2.1.1 目において取得した目的変数であるグラフト重合率、および、2.1.2 目において量子化学計算により算出した各種説明変数により構築した。機械学習では、データセット内に相関関係の高い説明変数ペアがある場合、回帰分析が不安定になる（多重共線性問題）ため、高相関関係にある説明変数ペアの一方の説明変数を予め除外するなどの前処理が必要となる。各説明変数クラス内における説明変数の相関プロットを作成したところ、いくつかの説明変数ペアにおいて、強い相関関係があることがわかった。そこで本研究では、説明変数間の相関係数の絶対値が 0.7 以上（ $|$ 相関係数 $| > 0.7$ ）の説明変数ペアのうち、グラフト重合率に対する相関係数が低い説明変数を除外することにより説明変数の削減を実施した。その結果、全 49 種類の説明変数のうち 28 種類の説明変数が残り、この 28 種類の説明変数を機械学習のためのデータセットとして採用した（Table 1, Highly correlated pair processing 参照）。28 種類の説明変数の内訳は、モノマーの物性情報が 3 種類（④、⑨

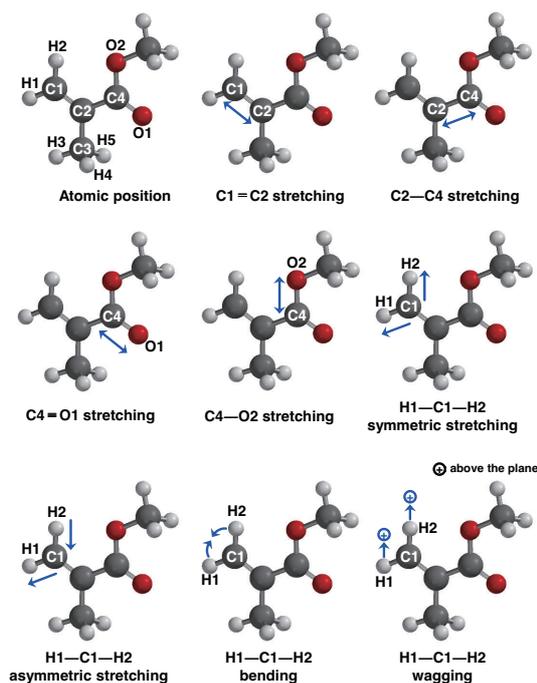


Figure 3. Atomic positions and FT-IR vibrational parameters for the methacrylate ester monomers. (Reprinted from Ref. 16, with the permission of Elsevier publishing.)

, ⑩), 原子電荷が 6 種類（⑭, ⑮, ⑰, ⑳, ㉑, ㉒）, NMR 化学シフトが 6 種類（㉔, ㉕, ㉖, ㉗, ㉘, ㉙）, そして FT-IR スペクトルが 13 種類（㉔, ㉕, ㉖, ㉗, ㉘, ㉙, ㉚, ㉛, ㉜, ㉝, ㉞, ㉟, ㊱, ㊲, ㊳, ㊴, ㊵, ㊶, ㊷, ㊸, ㊹）である。

2.3 機械学習

機械学習では、一般的に AI モデルの説明可能性とその予測性能はトレードオフの関係にあり、AI モデルの説明可能性の観点からブラックボックス型 AI モデルとホワイトボックス型 AI モデルに大別される。ブラックボックス型 AI モデルは、高精度な予測が可能であるが、その反面、学習過程が複雑であり、どの説明変数が予測結果に寄与するのかを人間が理解することは難しく、「理由は気にしないので、とにかく予測精度を高くしたい」という局面には適している。一方、ホワイトボックス型 AI モデルは、ブラックボックス型 AI モデルと比較すると予測精度は低い傾向にあるが、その反面、AI モデル間の説明変数と目的変数の関係を比較的容易に理解しやすいため、「説明変数の値が増減したときに、予測結果がどの程度増減するのか知りた

Table 1. Selection methods and results for explanatory variables.

Selection method of explanatory variable	Original	Highly correlated pair processing	AIC	BIC	LASSO
Total parameters of explanatory variables	49	28	24	18	26
Basic property	11	3	3	3	3
① Molecular weight	✓				
② E _{HOMO}	✓				
③ E _{LUMO}	✓				
④ Dipole moment	✓	✓	✓	✓	✓
⑤ Molecular surface area	✓				
⑥ Molecular volume	✓				
⑦ Ovality	✓				
⑧ Log P value	✓				
⑨ Polarizability	✓	✓	✓	✓	✓
⑩ Electronegativity	✓				
⑪ Hardness	✓	✓	✓	✓	✓
Atomic charge	11	6	4	3	6
⑫ C1 NBO charge	✓				
⑬ C2 NBO charge	✓				
⑭ C3 NBO charge	✓	✓	✓		✓
⑮ C4 NBO charge	✓	✓	✓	✓	✓
⑯ H1 NBO charge	✓				
⑰ H2 NBO charge	✓	✓			✓
⑱ H3 NBO charge	✓				
⑲ H4 NBO charge	✓				
⑳ H5 NBO charge	✓	✓	✓	✓	✓
㉑ O1 NBO charge	✓	✓			✓
㉒ O2 NBO charge	✓	✓	✓	✓	✓
Atomic NMR shift	11	6	6	5	5
㉓ C1 NMR shift	✓				
㉔ C2 NMR shift	✓	✓	✓	✓	✓
㉕ C3 NMR shift	✓	✓	✓		✓
㉖ C4 NMR shift	✓	✓	✓	✓	✓
㉗ H1 NMR shift	✓	✓	✓	✓	
㉘ H2 NMR shift	✓				
㉙ H3 NMR shift	✓				
㉚ H4 NMR shift	✓				
㉛ H5 NMR shift	✓				
㉜ O1 NMR shift	✓	✓	✓	✓	✓
㉝ O2 NMR shift	✓	✓	✓	✓	✓
FT-IR frequency and intensity	16	13	11	7	12
㉞ $\nu_{C1=C2, str.}$	✓	✓	✓	✓	✓
㉟ $i_{C1=C2, str.}$	✓	✓			✓
㊱ $\nu_{C2-C4, str.}$	✓	✓	✓		✓
㊲ $i_{C2-C4, str.}$	✓	✓	✓	✓	✓
㊳ $\nu_{C4=O1, str.}$	✓				
㊴ $i_{C4=O1, tr.}$	✓	✓	✓	✓	✓
㊵ $\nu_{C4-O2, str.}$	✓	✓	✓	✓	✓
㊶ $i_{C4-O2, str.}$	✓	✓	✓		✓
㊷ $\nu_{H1-C1-H2, sym. str.}$	✓	✓	✓	✓	✓
㊸ $i_{H1-C1-H2, sym. str.}$	✓				
㊹ $\nu_{H1-C1-H2, asym. str.}$	✓	✓	✓		✓
㊺ $i_{H1-C1-H2, asym. str.}$	✓				
㊻ $\nu_{H1-C1-H2, ben.}$	✓	✓			
㊼ $i_{H1-C1-H2, ben.}$	✓	✓	✓	✓	✓
㊽ $\nu_{H1-C1-H2, wag.}$	✓	✓	✓	✓	✓
㊾ $i_{H1-C1-H2, wag.}$	✓	✓	✓		✓

い、または、どの説明変数が予測結果に重要な因子であるのかを知りたい」という局面に適している。本研究では、作成した AI モデルの化学的妥当性を検証するといった観点から、ホワイトボックス型 AI モデルの作成を試みた。

機械学習には、統計解析向けのプログラミング言語である「R 言語」を使用した。グラフト重合率の予測モデル (AI モデル) の作成には、200 種類以上の機械学習手法を簡単なプログラミングで扱える caret パッケージを利用した。本研究では機械学習手法として、ホワイトボックス型 AI モデルが作成可能な線形重回帰 (multiple linear regression), 決定木 (decision tree), および、2 種類のアンサンブル学習 (ランダムフォレスト (random forest) と勾配ブースティング (extreme gradient boosting)) を採用した。

線形重回帰は、作成した AI モデルが数式化されているため視覚的に理解しやすいといったメリットがあるが、トレーニングデータに外れ値が含まれていると予測精度が著しく低下するといったデメリットがある。決定木は、ある閾値を基準にデータを非線形に分割・分析することが可能な手法であり、外れ値を含むようなデータに対して有効であるが、決定木単体 (1 本の決定木) では比較的低い予測精度が低い傾向にある。この決定木の予測精度が低いといった欠点を補う手法として、アンサンブル学習がある。アンサンブル学習のひとつであるランダムフォレストは、ランダムに抽出したトレーニングデータから数百~数千の少しずつ異なる決定木を作り出し、それら決定木の予測結果を平均化して、最も高かったものを最終的に AI モデルとして出力する手法である。勾配ブースティングは、最初に 1 本の決定木を作り、その決定木の誤りを数百~数千回逐次的に修正して、最終的な AI モデルを出力する手法である。ランダムフォレスト、勾配ブースティングのいずれの手法においても、作成される AI モデルは非線形モデルであるため数式化できないといったデメリットがあるが、トレーニングデータ内に外れ値が含まれていても安定的に高い予測精度を維持できる (堅牢なモデル) といったメリットがある。

本研究では、線形重回帰には caret パッケージに実装されているアルゴリズム「glmnet」を、決定木には「rpart2」、ランダムフォレストには「rf」、勾配ブースティングには「xgbTree」を使用した。各 AI モデルの挙動を設定・制御するためのハイパーパラメータ探索は 10 分割交差検証により実施した。作成した AI モデルの予測精度は、決定係数 (coefficient of determination, R^2) と二乗平均平方根誤差 (root mean square error, RMSE)

により評価した。

3 結果と考察

3.1 線形重回帰によるグラフト重合率予測モデルの作成

AI モデルの予測精度は、機械学習に用いるアルゴリズムの種類により大きく変化することが知られている。ここでは、線形重回帰により予測モデルを作成した結果を示す。2.2 項で選定した 28 種類の説明変数を用いて線形重回帰モデルを作成した結果を Fig. 4A に示す。青プロットがトレーニングデータの結果であり、赤プロットがテストデータの結果である。X 軸は AI モデルにより算出した予測グラフト重合率であり、Y 軸は実測により得た実測グラフト重合率である。グラフ上の破線上にデータがプロットされた場合、予測値と実測値が完全に一致していることを意味し、破線から離れるほど予測誤差が大きいのことを意味する。

Figure 4A に示すように、全 28 種類の説明変数を用いた線形重回帰モデルは、トレーニングデータとテストデータともに予測誤差が大きく、十分な予測精度を持つ AI モデルではなかった。特に、未知モノマーに対する予測誤差が大きくなり、汎化性能 (未知データに対応する能力) が低い AI モデルであった。この低汎化性能の原因としては、説明変数 28 種類すべてを使用することにより、グラフト重合反応に対する寄与度が少ない説明変数が AI モデルに組み込まれたためであると考えられる。そこで、線形重回帰モデルの汎化性能向上を目指し、寄与度が少ない説明変数を除外した AI モデルの作成を試みた。

線形重回帰では、より多くの説明変数を用いて作成された多項式 AI モデルは、より複雑な関数となり、トレーニングデータによく適合する傾向にあるが、その一方で未知モノマーの予測精度は低くなり、また、AI モデルの説明・解釈も複雑化する。AI モデル説明性の観点から、より簡単な AI モデルが好まれることは一般的にも当然である。しかしながら、説明変数の取捨選択を研究者自身が行うと、恣意的な説明変数選択となる可能性がある。

このような問題を解決する手法として、統計学的に説明変数を取捨選択する赤池情報量規準 (Akaike's Information Criterion, AIC) やベイズ情報量規準 (Bayesian information criterion, BIC), あるいは、数理的な正則化手法により説明変数選択をする L1 正則化 (least absolute shrinkage and selection operator, LASSO) がある。本稿では、AIC, BIC, LASSO については詳細に解説

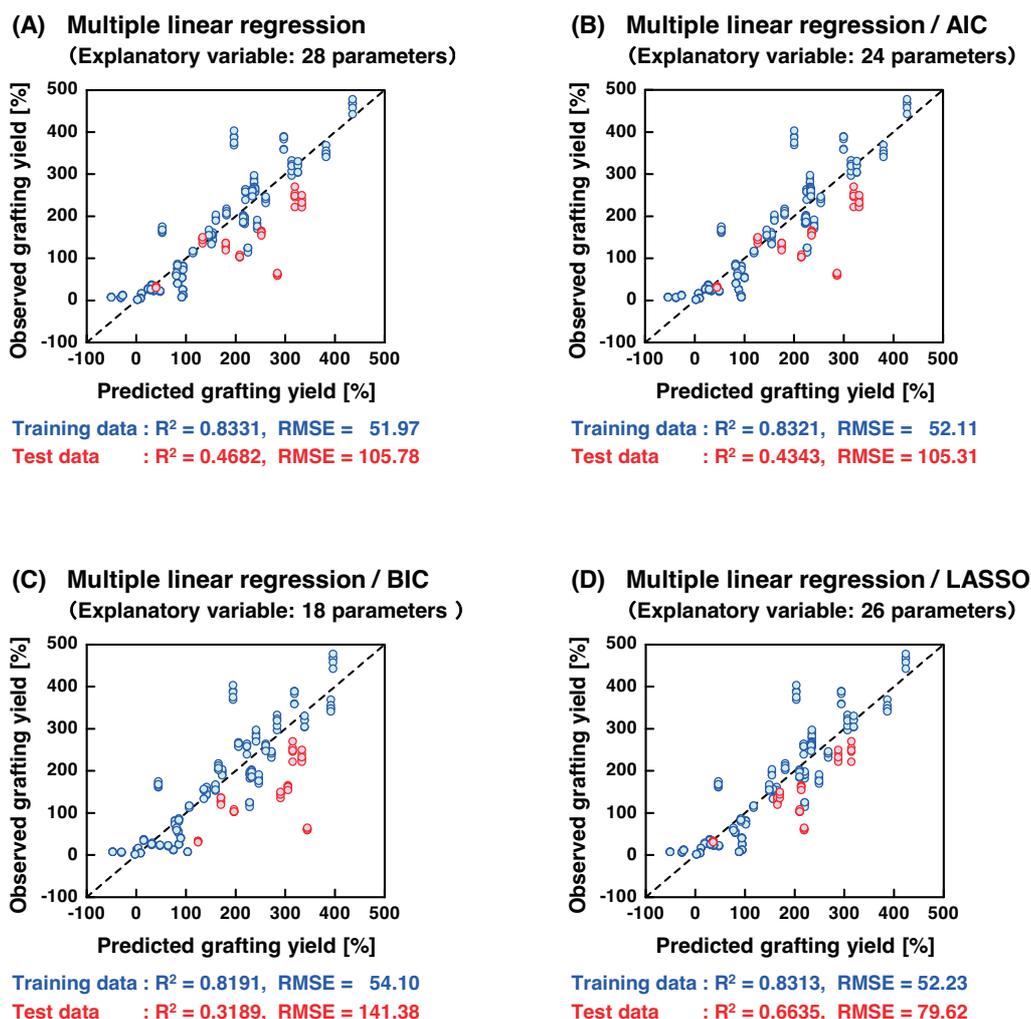


Figure 4. Comparison of explanatory variable selection criteria in multiple linear regression. (A) Multiple linear regression without explanatory variable selection criteria. (B) Multiple linear regression with AIC. (C) Multiple linear regression with BIC. (D) Multiple linear regression with LASSO. (Reprinted from Ref. 16, with the permission of Elsevier publishing.)

しないが、いずれの手法においても「モデルの複雑さと、データとの適合度とのバランスを取る」ように数理的判断規準により説明変数が自動的に選択されるため、恣意性のない AI モデルが作成可能となる。

さらにいえば、アルゴリズムの種類と同様に、説明変数選択手法の違いにより AI モデルの予測精度は変化するため、本研究では線形重回帰における説明変数選択手法の効果についても検討した。Table 4A に AIC, BIC, LASSO による説明変数選択結果を、Fig. 4B, C, D に各種説明変数選択処理後の AI モデルによるグラ

フト重合率の予測結果を示す。各種説明変数処理を実施した結果、線形重回帰/AIC モデルの説明変数は 24 種類（除外説明変数：4 種類：⑰, ⑳, ㉓, ㉔）となった。線形重回帰/BIC モデルの説明変数は 18 種類（除外説明変数：10 種類：⑭, ⑰, ⑳, ㉓, ㉔, ㉕, ㉖, ㉗, ㉘, ㉙）となった。そして、線形重回帰/LASSO モデルの説明変数は 26 種類（除外説明変数：⑳, ㉔）となった。

Figure 4 に示すように、線形重回帰/AIC モデルと線形重回帰/BIC モデルでは、説明変数 28 種類を用いた

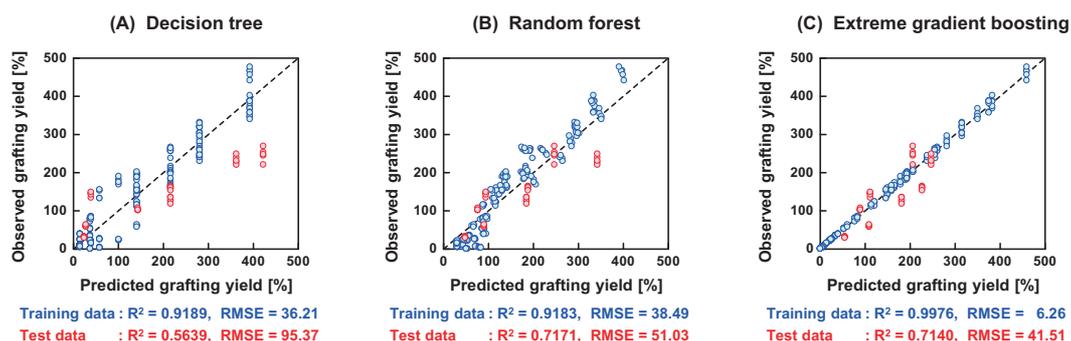


Figure 5. Predicted versus observed grafting yields plot; (A) decision tree, (B) random forest, and (C) extreme gradient boosting. (Reprinted some modifications from Ref. 16, with the permission of Elsevier publishing.)

線形重回帰モデルと同様に予測誤差が大きく、予測精度の向上を図ることができなかった。一方、線形重回帰/LASSOモデルでは、テストデータに対する二乗平均平方根誤差 (RMSE = 79.62) が大きく改善され、AIモデルの汎化性能が向上した。しかしながら、線形重回帰/LASSOモデルにおいても依然として予測誤差は大きく、十分な予測AIモデルではないことがわかった。

3.2 決定木、および、アンサンブル学習によるグラフト重合率予測モデルの作成

3.1項の結果から、グラフト重合率予測において線形重回帰では予測精度の高いAIモデルが作成できないことがわかった。そこで次に、非線形分析に対応可能な決定木、および、2種類のアンサンブル学習 (ランダムフォレストと勾配ブースティング) によりグラフト重合率予測モデルの作成を試みた。ここでは特別な説明変数選択処理は実施せずに、2.2項で選定した28種類の説明変数を用いて各種AIモデルを作成した (Table 1, Highly correlated pair processing 参照)。

Figure 5Aは決定木モデルの予測結果であり、予想通りに決定木モデルの予測精度は低い値となった。一方、アンサンブル学習であるランダムフォレスト (Fig. 5B) と勾配ブースティング (Fig. 5C) の予測結果は、線形重回帰モデルと比較すると、予測誤差が低くなり、高精度なグラフト重合率予測を実現できることがわかった。特に勾配ブースティングモデルにおいては、反応成績の知られていない未知モノマーの物性情報を入力するだけで、高い精度でそのグラフト重合率を瞬時に予測することに成功した (テストデータ:

$R^2 = 0.71$, RMSE = 41.51)。この際の未知モノマーに対するRMSE値は、線形重回帰/LASSOモデルの約半分の値となり、AIモデルの汎化性能は大幅に改善された。これは、アンサンブル学習の「外れ値やノイズに対して影響を受け難く、安定な予測ができる」といった特徴を反映した結果に由来するものと推察される。つまり、本研究ではトレーニングデータに用いたモノマー数が41種類と少ないスモールデータ分析であるため、データの偏りや外れ値が発生しやすく、故に、外れ値の影響を受けやすい線形重回帰分析よりも、より外れ値耐性の高いアンサンブル学習が最適な学習アルゴリズムとして選定されたものと推察される。

機械学習では作成したAIモデルを解析することにより、どの説明変数がグラフト重合率を支配する重要な反応因子となるのかを客観的に評価することができる。そこで、作成した勾配ブースティングAIモデルを解析し、予測グラフト重合率に及ぼすモノマーの物性情報の影響度を検討した。Figure 6には、勾配ブースティングAIモデルにおける重要説明変数トップ10を示す。ここでは、相対重要度の数値が大きい説明変数ほど、より重要な説明変数であることを意味する。

Figure 6に示すようにAIモデルを構成する28種類の説明変数のうち、「⑨モノマーの分極率」と、モノマーの置換基近傍にある「③酸素原子 (O2) のNMR化学シフト」がグラフト重合反応の鍵となる重要説明変数であることを発見した。一方、機械学習は入力データに潜む規則性を見つけ出すものであり、重要説明変数として選択された説明変数に対しては、別途、化学的にその妥当性を検討する必要がある。説明変数「⑨モノマーの分極率」は、モノマーの分子構造全体に

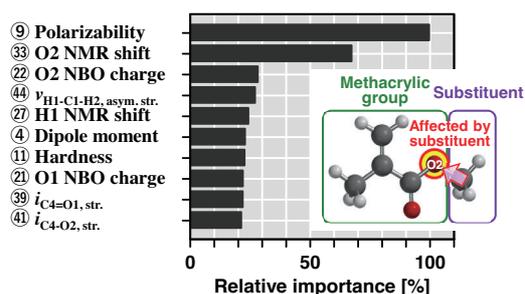


Figure 6. The top 10 most important explanatory variables in the predictive Extreme gradient boosting model. (Reprinted some modifications from Ref. 16, with the permission of Elsevier publishing.)

起因する説明変数であり、分子構造の異なるモノマーを対象としてグラフト重合率予測モデルを作成している本研究においては、「⑨モノマーの分極率」が最重要説明変数として選択されることは化学的な解釈からも妥当な結果である。

一方、説明変数「③③酸素原子 (O2) の NMR 化学シフト」が 2 番目に重要度の高い説明変数として選択された要因は、置換基の影響が「③③酸素原子 (O2) の NMR 化学シフト」に反映されたためであると考察できる。つまり、置換基の最近傍に存在する「酸素原子 (O2)」は、他の原子よりも置換基の影響を受けやすい状態にあるため、置換基構造の違いが酸素原子の物理的・化学的な特性変化を顕在化させたものと推察される。このように、置換基最近傍に存在する酸素原子が置換基の影響を受けやすいといった考察は、一般的な化学的解釈と良く一致しており、これもまた化学的に妥当性のある説明変数選定であると理解できる。特に、この「酸素原子の NMR 化学シフト」は、これまで研究者が見逃していた有益な科学的知見であり、まさに機械学習の活用なしでは発見できなかった結果である。このような機械学習の利活用が「新たな知の創造」や「基礎研究の深化・発展」に繋がるものと期待できる。

4 まとめと今後の展望

本稿では、高分子材料開発における機械学習の活用例として、グラフト重合反応に使用するモノマーの物性情報のみからグラフト重合率を瞬時に予測可能な AI モデル作成について紹介した。本研究では、トレーニングデータに用いたモノマー数が 41 種類と少

ないスモールデータ分析であったものの、アンサンブル学習のひとつである勾配ブースティングを適用することにより未知モノマー (テストデータ) に対して高い予測精度を示す AI モデルを作成することに成功した。また、作成した AI モデルの解析により、これまで研究者が着目していない説明変数である「酸素原子の NMR 化学シフト」がグラフト重合反応に寄与していることを見出すことができた。これらの結果は、放射線グラフト重合に対する機械学習の有用性を示すものである。今回創出した AI モデルを活用、並びに、アップデートすることにより、試行錯誤をとまなう繰り返し実験は不要となり、数年単位が当然だった材料開発の短期化とそれにとまなう開発コストの削減が実現できるものと期待される。今後は、異なる分子構造のモノマーの物性情報を AI モデルに学習させることによる予測精度の向上と、放射線重合だけでなく広範な高分子重合反応への適用を図るとともに、材料合成に必須な放射線の照射線量、モノマー濃度、重合反応温度、重合反応時間などの諸条件を考慮した AI モデルの作成に展開し、最適なグラフト重合反応の合成条件の提案を可能とする技術の構築に繋がりたいと考えている。

〈謝 辞〉

本研究は JSPS 科研費 JP20K12488 および QST アライアンス事業「先端高分子機能性材料アライアンス」の一環として実施したものです。

〈参 考 文 献〉

- 1) F. Strieth-Kalthoff, F. Sandfort, M. H. S. Segler, F. Glorius, *Chem. Soc. Rev.*, 49 (2020) 6154.
- 2) Y. Mo, S. P. Ong, G. Ceder, *Chem. Mater.*, 24 (2012) 15.
- 3) M. Nishijima, T. Ootani, Y. Kamimura, T. Sueki, S. Esaki, S. Murai, K. Fujita, K. Tanaka, K. Ohira, Y. Koyama, I. Tanaka, *Nat. Commun.*, 5 (2014) 4553.
- 4) A. Seko, A. Togo, H. Hayashi, K. Tsuda, L. Chaput, I. Tanaka, *Phys. Rev. Lett.*, 115 (2015) 205901.
- 5) F. Ren, L. Ward, T. Williams, K. J. Laws, C. Wolverton, J. Hattrick-Simpers, A. Mehta, *Sci. Adv.*, 4 (2018) eaaq1566.
- 6) Z. Hou, Y. Takagiwa, Y. Shinohara, Y. Xu, K. Tsuda, *ACS Appl. Mater. Interfaces*, 11 (2019) 11545.
- 7) J. M. Granda, L. Donina, V. Dragone, D. -L. Long, L. Cronin, *Nature*, 559 (2018) 377.

- 8) K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis, M. Z. Bazant, S. J. Harris, W. C. Chueh, R. D. Braatz, *Nat. Energy*, 4 (2019) 383.
- 9) A. Yada, K. Nagata, Y. Ando, T. Matsumura, S. Ichinoseki, K. Sato, *Chem. Lett.*, 47 (2018) 284.
- 10) K. Ito, Y. Obuchi, E. Chikayama, Y. Date, J. Kikuchi, *Chem. Sci.*, 9 (2018) 8213.
- 11) M. Tamada, *Radiation Processing of Polymers and Its Applications*, in: H. Kudo (Eds.), *Radiation Applications*, Springer, Singapore, 2018, pp. 63–80.
- 12) 斎藤 恭一, 藤原 邦夫, 須郷 高信, *グラフト重合による吸着材開発の物語*, 丸善出版, 2019.
- 13) Y. Ueki, N. Seko, *ACS Omega*, 5 (2020) 2947.
- 14) N. Seko, H. Hoshina, N. Kasai, T. Shibata, S. Saiki, Y. Ueki, *Radiat. Phys. Chem.*, 143 (2018) 33.
- 15) A. M. A. Mahmoud, K. Yoshimura, Y. Maekawa, *J. Membr. Sci.*, 620 (2021) 118844.
- 16) Y. Ueki, N. Seko, Y. Maekawa, *Appl. Mater. Today*, 25 (2021) 101158.

〈著者略歴〉

植木 悠二：2006年群馬大学大学院工学研究科博士後期課程修了。2006年日本原子力研究開発機構（現量子科学技術研究開発機構）入社。2022年より上席研究員（現職）。博士（工学）。専門は放射線高分子化学。

瀬古 典明：1996年神奈川大学大学院工学研究科修士課程修了。1996年日本原子力研究所（現量子科学技術研究開発機構）入所。2010年よりグループリーダー、2016年よりプロジェクトリーダー（現職）。博士（工学）。専門は放射線高分子化学。

前川 康成：1991年東京大学院工系研究科合成化学専攻博士課程修了。1991年–1992年IBM Almaden Research Center 博士研究員。1992年–1993年University of Wisconsin (Madison) 博士研究員。1994年–1998年日立製作所日立研究所 研究員。1998年日本原子力研究所（現量子科学技術研究開発機構）入所。2022年より同研究所 所長（現職）。博士（工学）。専門は放射線化学、高分子化学。