

令和 6 年 6 月 7 日現在

機関番号：82502

研究種目：基盤研究(C)（一般）

研究期間：2020～2023

課題番号：20K12488

研究課題名（和文）放射線グラフト重合における機械学習手法の活用と重合収率予測モデルの構築

研究課題名（英文）Utilization of machine learning for radiation graft polymerization and construction of polymerization yield prediction model

研究代表者

植木 悠二（Ueki, Yuji）

国立研究開発法人量子科学技術研究開発機構・高崎量子応用研究所 先端機能材料研究部・併任

研究者番号：50446415

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：従来の高分子材料開発では、研究者の「経験と勘」に基づいた非効率的な試行錯誤的実験に頼ることが多く、新材料が完成するまでには膨大な時間と費用を必要とする。本研究では、高分子材料の改質手法である放射線グラフト重合技術において、従来の経験的な実験科学に機械学習手法を融合させることにより、重合反応に使用するモノマー（薬品）の物性情報のみからグラフト重合反応率を瞬時に予測できるAIモデルの創出に成功した。また、開発したAIモデルを解析したところ、モノマーの「分極率」と、モノマーの置換基近傍にある酸素原子の「NMR化学シフト」が重合反応予測の鍵となる重要因子であることが判明した。

研究成果の学術的意義や社会的意義

本成果は、機能性高分子材料の創製手法のひとつである放射線グラフト重合における機械学習利用の有用性を示したものである。本成果は、低コストで迅速性のある効率的な高分子材料開発に資する基礎技術であり、企業競争力向上に貢献可能であることから、その社会的意義は大きい。また、本成果の応用・発展は、高分子材料開発分野における新たな科学的知見の発見や革新的高分子材料の創出に繋がる可能性を有していることから、その学術的意義も大きい。

研究成果の概要（英文）：Conventional polymer material development relies on inefficient trial-and-error experiments based on researcher's "experience and intuition". As a result, the development of new polymers requires an enormous amount of time and high costs. In this research, we have succeeded in creating an AI model that can instantly predict the grafting yields based solely on the physical and chemical properties of the monomers, by integrating machine learning approach in the conventional radiation grafting process. Additionally, the creating AI model can quantify the importance of various explanatory variables on the grafting yield. Analysis of the AI model revealed that the monomer's "polarizability", which represents a miscibility indicator of the monomer to the trunk polymer, and the "O2 NMR shift", which represents a diffusivity indicator of the monomer into the trunk polymer, were important explanatory variables for predicting the grafting yield.

研究分野：量子ビーム科学関連

キーワード：量子ビーム 放射線グラフト重合 機械学習 重合予測

1. 研究開始当初の背景

現代社会の発展は、プラスチックの「進化」により躍進してきた。しかし、従来のプラスチック材料の研究開発では、研究者の「経験と勘」に基づいて仮説を立て、試行錯誤的実験を繰り返すことで最適な構造や組成を求めてきた。そのため、新材料の創製には、膨大な時間と高いコストが掛かることが問題であった。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）を経験や勘だけを頼りに見つけ出すことは、ますます困難な状況となりつつある。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、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、短時間かつ低コストで実験結果の事前予測や最適解の提案が可能となる。さらに、AI の活用により、これまで研究者が見逃していた有益な科学的知見や新規材料の発見に繋がっており、世界中で AI の利活用が精力的に進められている。

研究代表者は、高分子材料加工技術のひとつである放射線グラフト重合技術を駆使し、これまでに超強塩基性領域において機能する金属吸着材や、環境水に含まれる放射性セシウム除去材の開発に成功してきたものの、その研究開発期間は長期間を要した。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の維持・強化のためには、放射線グラフト重合研究においても AI や MI の利活用は必須である。そこで研究代表者は、放射線グラフト重合技術に機械学習を融合させることにより、薬品の分子情報を元にグラフト重合収率を高精度に予測可能な解析手法を開発し、開発期間短縮とコスト削減に資する基盤技術の確立を目指した。

2. 研究の目的

本研究では、放射線グラフト重合技術への機械学習利用の有用性検証として、薬品などの分子情報のみからグラフト重合収率を予測可能な AI モデルの作成、並びに、作成した AI モデル解析によるグラフト重合反応における重要支配因子の解明を目的とする。

3. 研究の方法

3.1. データ収集

図1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、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ショングラフト重合反応をモデル反応として設定した。モノマーに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種類を学習データ、残り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収率は、電子線エマルショングラフト重合反応実験により取得した。また、説明変数として用いるモノマーの物性情報は、量子化学計算（計算ソフト：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。説明変数としては、モノマーの分子量や体積などの基礎物性情報に加えて、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種類、水素原子：5種類、酸素原子：2個）の電荷やNMR化学シフト、また、メタクリル酸基の赤外吸収振動波数とその振動強度など合計49種類の変数を用意し、データセットとした。

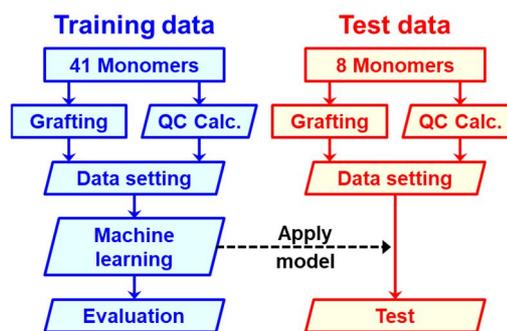


図1. 機械学習の分析フローチャート

3.2. 機械学習

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必要がある。本研究におけるデータセットには、高相関関係にある説明変数ペアが存在していた。そこで、説明変数間の相関係数の絶対値が0.7以上（ $|\text{相関係数}| > 0.7$ ）の説明変数ペアのうち、グラフト重合収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。

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、グラフト重合収率予測モデル（AIモデル）の作成を試みた。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、決定木、及び、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(XGBoost))を採用した。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。

4. 研究成果

4.1. グラフト重合収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、各種アルゴリズムにより予測モデルを作成した結果を図 2 に示す。青プロットが学習データの結果、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収率、Y 軸は実測により得た実測グラフト重合収率を示す。グラフ上の破線の上にデータがプロットされれば、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ことを意味する。

図 2 (A) に示す線形重回帰モデルは、訓練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。そこで、線形重回帰モデルの汎化性能向上を目指し、赤池情報量規準 (AIC) やベイズ情報量規準 (BIC)、L1 正則化 (LASSO) を適用して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。しかし、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分かった (訓練データ: $R^2 = 0.8313$, RMSE = 52.23, テストデータ: $R^2 = 0.6635$, RMSE = 79.62)。

次に、非線形分析に対応可能な決定木、及び、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収率予測モデルの作成を試みた。図 2 (B) 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 (図 2 (C)) と勾配ブースティング (図 2 (D)) 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収率予測を実現できることが分かった。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収率を瞬時に予測することに成功した (XGBoost/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、本研究では訓練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。

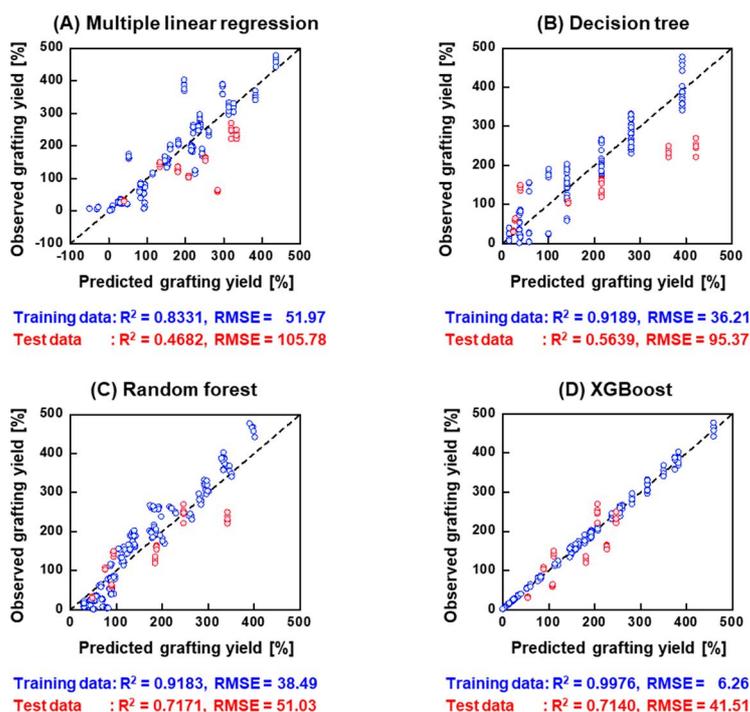


図 2. 各種アルゴリズムにおけるグラフト重合収率予測結果

A: 線形重回帰, B: 決定木, C: ランダムフォレスト, D: XGBoost

4.2. グラフト重合収率予測モデルの作成

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収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、予測グラフト重合収率に及ぼすモノマーの物性情報の影響度を検討した。図 3 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。図 3 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、「モノマーの分極率」と、モノマーの置換基近傍にある「³³ 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「モノマーの分極率」は、モノマーの分子構造全体に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収率予測モデルを作成している本研究においては、「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。

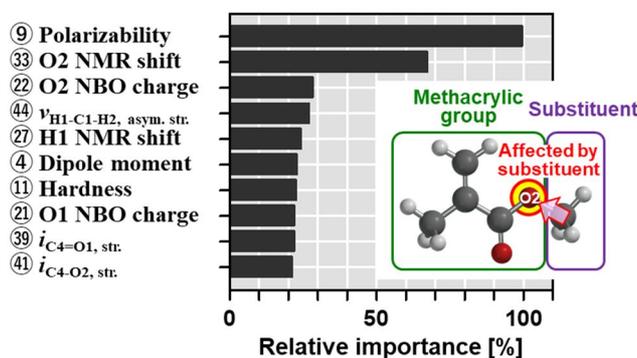


図 3. XGBoost モデルにおける重要説明変数トップ 10

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 植木悠二, 瀬古典明, 前川康成	4. 巻 114
2. 論文標題 機械学習を活用した放射線グラフト重合率の予測	5. 発行年 2022年
3. 雑誌名 放射線化学	6. 最初と最後の頁 45-54
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Yuji Ueki, Noriaki Seko, Yasunari Maekawa	4. 巻 QST-M-39
2. 論文標題 Prediction of grafting yield by using machine learning	5. 発行年 2023年
3. 雑誌名 QST Takasaki Annual Report 2021	6. 最初と最後の頁 51-51
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Ueki Yuji, Seko Noriaki, Maekawa Yasunari	4. 巻 25
2. 論文標題 Machine learning approach for prediction of the grafting yield in radiation-induced graft polymerization	5. 発行年 2021年
3. 雑誌名 Applied Materials Today	6. 最初と最後の頁 101158-101158
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.apmt.2021.101158	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Ueki Yuji, Seko Noriaki, Maekawa Yasunari	4. 巻 QST-M-33
2. 論文標題 Prediction of grafting yield by multiple linear regression analysis	5. 発行年 2021年
3. 雑誌名 QST Takasaki Annual Report 2020	6. 最初と最後の頁 50-50
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計3件（うち招待講演 1件 / うち国際学会 0件）

1. 発表者名 植木悠二
2. 発表標題 機械学習による放射線グラフト重合収率予測
3. 学会等名 QST高崎サイエンスフェスタ2022
4. 発表年 2022年

1. 発表者名 Yuji Ueki
2. 発表標題 Prediction of radiation-induced graft polymerization yield by using machine learning
3. 学会等名 Joint Symposium of S-Membrane Project and F-Material Project (招待講演)
4. 発表年 2022年

1. 発表者名 植木悠二
2. 発表標題 人工知能 (AI) でグラフト重合反応率が予測可能に
3. 学会等名 QST高崎サイエンスフェスタ2021
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

どの原料モノマーを使えば、どんな高分子材料を作れるか分かる！？（プレスリリース）
<https://www.qst.go.jp/site/press/20210929-1.html>

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------