

LLMとMIで革新する 素材開発プラットフォーム

小淵 喜一 船矢 幸一 當山 清彦 田中 修吉 ダニエル オノロルビオ マーティン レンチェン ミン イアン メルビン

要 旨

本稿では、素材開発分野への大規模言語モデル（LLM）応用の取り組みを紹介します。NECでは素材開発プラットフォームの開発を進めています。調査活動、実験計画の2つの素材開発ステップに対して、それぞれ対応するコア技術を適用することで、論文・レポートなどの文書や実験結果などのデータを整理してユーザーにAIが対話的に提示したり、機械学習のモデルに物理や化学の原理を反映させる手法により、少ないデータでも学習でき正確な物性予測をすることが可能となります。このプラットフォームにより、マテリアルズ・インフォマティクス（MI）技術と膨大な業界文献・知識のシームレスな統合を実現し、素材開発プロセスに革新をもたらすことを目指しています。



素材領域特化型LLM／知識グラフ／オントロジー／情報抽出技術／マテリアルズ・インフォマティクス（MI）／PIML（Physics Informed Machine Learning）／分子構造の生成AIモデル

1. はじめに

工業用素材に対する高性能化やニーズの多様化、更には環境への配慮が求められるなか、素材開発においては効率的な開発が求められています。現在、素材開発者は、大きく分けて（1）調査活動：文献や特許、社内技術レポートやお客様ニーズの調査活動、及び（2）試作検討：実験計画を基にした実験の繰り返し、といった2つの作業に取り組みますが、膨大な資料から欲しい情報を得ることや、広大な探索空間のなかから欲しい素材設計を探し当てることは容易ではなく、どうしても時間を浪費してしまうという課題があります。

そのなかでも高分子素材は、身の回りのあらゆる箇所で使用されている素材であり、さまざまな機能を満たす素材を素早く開発することが求められています。しかし、膨大な過去の社内資料が死蔵されていて、調査が満足にできないケースも散見します。また、試作検討についても、大規模なオープンデータベースがなく、シミュレーションやハイスループット実験（実験自動化）もハードルが高く、十分なデータを得られないといった課題も存在しています。NECでは大規模言語モデル（以下、LLM）をはじめとする生成AI技術を素材開発分野に応用して、これらの課題を解決する素材

開発プラットフォームの開発に取り組んでいます。

まず、図1のように、このプラットフォームを活用することで、死蔵されていた社内データや論文・特許・オープンデータベースといった社外の文書の活用が可能になります。その文書から独自の技術で効率的に素材に関する知識をLLMが学習し、それを調査活動の支援に生かしたり、抽出されたデータをNEC独自のモデリング技術を使って

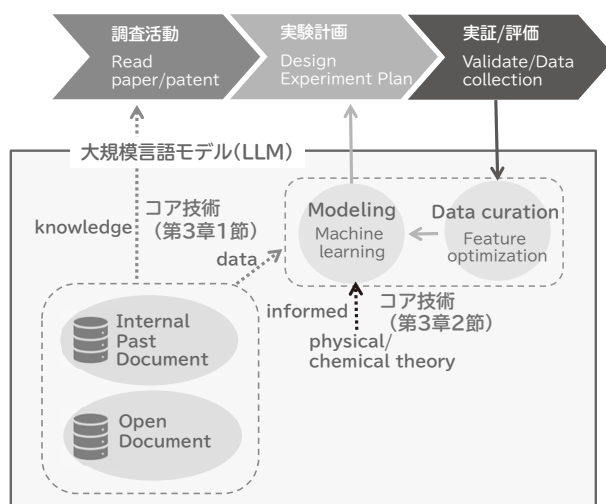


図1 素材開発プラットフォーム

物性予測や実験計画に活用したりします。

また、LLMを通じたチャット方式によりAI活用のハードルを大幅に下げます。より具体的には、今までのAI活用において障壁となっていたプログラミングの習得やツールの使いこなしの必要性をチャットという気軽な利用方式を導入することで解消し、デジタルに対して苦手意識を持つ方も含めて、多くの素材開発者が本業である素材開発に時間を振り向けられるようにする、またそれが「素材開発プラットフォーム」に搭載されたAIによって更に高度化する、といったことを狙います。

本稿では、この素材開発プラットフォームについて概説したのち、そこに搭載されるAIのコア技術を説明します。

最後に、実際の適用事例として、バイオプラスチックでの事例を紹介します。

2. 素材開発プラットフォームとは？

図2に素材開発プラットフォームの概略を示します。

この素材開発プラットフォームでは、素材業界において材料開発の作業に掛かっている手間と時間を削減して効率化することを目指します。そのため生成AIをはじめとする先端技術を活用することで、図1に示した作業手順の各々の段階での作業を次のように効率化します。

まず、調査活動の段階では、論文・レポートなどの文書や実験結果などのデータを整理してユーザーに対話的に提示することで、膨大な情報ソースから開発対象となる素材の分子構造候補を簡単に検索して絞り込んでいくことを可能にします。

次に実験計画の段階では、少ないデータでも学習ができ、正確な物性予測をすることができる仕組みを提供します。後述しますが、この仕組みの実現にあたっては、機械学習のモデルに物理や化学の原理を反映させる、という手法を採用しています。このようにして開発された機械学

習モデルを利用することで、必要な物性を実現する可能性が高いと判断される分子構造を知ることができ、実験検証の対象となる分子構造を絞ることで従来膨大な作業を要する実験評価の段階を効率化することにつながります。

第3章では、このような素材開発プラットフォームの実現で重要となるコア技術を紹介します。

3. 素材開発プラットフォームにおけるコア技術

3.1 素材候補の背景情報検索・絞り込み

素材開発プラットフォームでの最初の作業は、候補となる素材に関連する背景情報を幅広く集め、開発作業の方針や素材候補の目処をつけながら絞り込んでいくことから始まります。図2の「調査活動」がそれにあたります。

素材開発プラットフォームでは、この作業を次のステップで実現します。

- 1) 情報ソースから素材情報を自動に取り込む
- 2) 対話インターフェースを使って関連情報を検索・絞り込む

ステップ1) では論文や社内レポート、実験データなどアクセス可能な情報ソースから素材の特性や製造方法に関する情報をくまなく抽出し、いわゆる知識グラフのフォーマットでデータベースに格納します。知識グラフとは、主語、述語、目的語からなるタプル（＝データ単位）のリストで構成されるデータフォーマットです¹⁾²⁾。素材業界で利用しやすいよう、素材開発プラットフォームでは各々の要素（すなわち主語、述語、目的語）を標準的な物性科学のオントロジーで定義し、各要素間の関係を知識グラフ・スキーマで定義することで、素材開発の業務に耐えうる明確さや再現性を担保します。

更にLLM上で前述したオントロジーと知識グラフ・スキーマを深層学習させることで、複雑な言語処理パイプラインを必要とせずに文献からダイレクトに知識グラフを抽出する方法を開発しました。これにより、膨大な文献から簡単な手続きで正確かつ高速に情報を取り出すことが可能です。

前述の取り出された知識グラフを活用することで、背景情報の詳細な検索が可能となるだけでなく、生成AIの学習データを準備する手間も大幅に減らし、素材特性を予想するための予想AIを、あるいは欲しい素材特性を実現する分子構造の生成AIを学習させるための手間を削減し

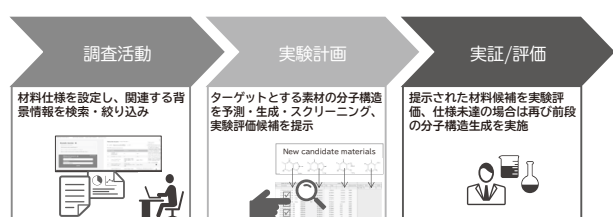


図2 素材開発プラットフォームの概略

ます(図3)。

ステップ2)では、ステップ1)で獲得した知識グラフと文献を合わせてLLM上で推論エンジンに学習させることで、対話的に分子構造の候補を絞る仕組みを実現しています。対話インタフェースでは、素材設計で重要となる複数の設計条件を対話的に入力することで、分子構造の候補群を取捨選択しながら絞る作業を簡単な操作で処理できるようにしています。例えば、まず開発したい高分子の機械・温度・電気特性や製造方法に関する条件を自然言

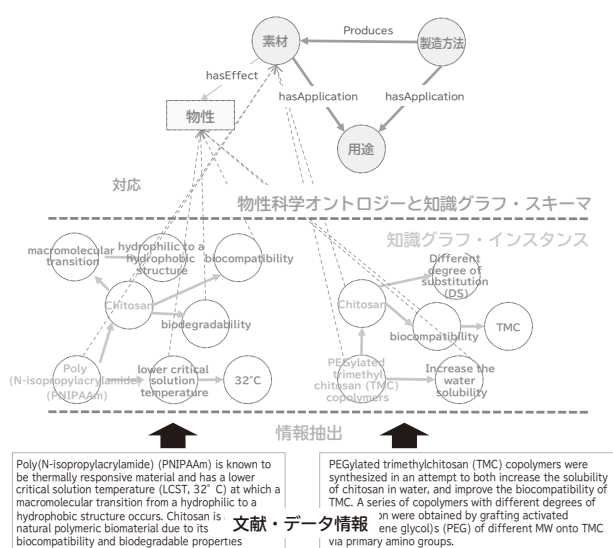


図3 ステップ1)の構造

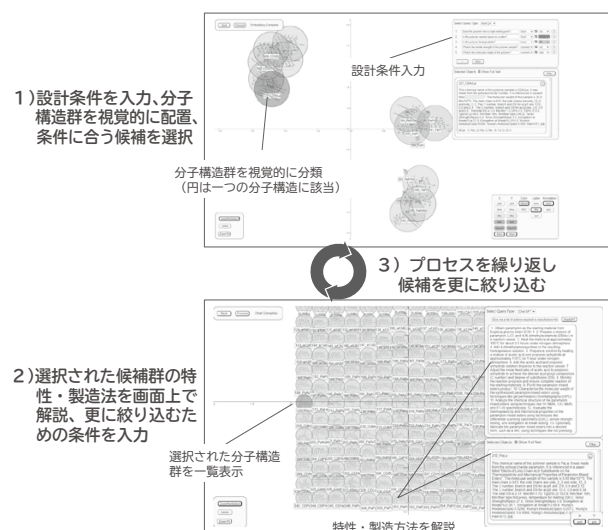


図4 ステップ2)の構造

語などで設定し、分子構造の各々の候補を条件への合致具合に応じて画面に配置、視覚的に目的に適いそうな分子構造を選択していきます。選択するたびに候補の特性や製造方法の解説が表示され、それを読んで選択操作が適切であることを確認できます(図4)。

3.2 高分子の特徴を組み込んだ機械学習モデルと分子生成技術

マテリアルズ・インフォマティクス(MI)は、AIなどの情報科学を活用して素材開発を支援する分野として、近年急速に発展しています。しかし、結晶構造と物性との関係性が深い無機材料や化学構造の表現がしやすい有機低分子に比べて、有機高分子は機械学習の応用が最も難しい素材の1つと言われてきました。長い鎖同士が絡まった高分子鎖の表現、更には混合物におけるフィラーや添加剤との相互関係性、ゴム素材における架橋など、単に高度な機械学習のアーキテクトを当てはめただけでは、応用できない技術的課題が存在してきたからです。

そこで、NECは、高分子の物理法則を機械学習モデルに組み込むアプローチ(PIML: Physics-Informed Machine Learning)でこれらの課題を解決する技術開発を行っています。具体的には、機械学習に入力する特徴量に物理や化学の原理から得られる値を使う手法や、高分子特有の特徴を表す理論式を機械学習アーキテクトに組み込む技術の構築を行っています。また、分子自体を生成することにも取り組んでおり、高分子の専門知識に基づいたルールベースと生成AIを適用したモデルベースの両方の良さを追求しながら技術開発を進めています。

このように、PIML及び分子生成技術によって、広大な探索空間に対してデータが少なく機械学習モデルが機能しないという現状を打破できるような革新的な技術の開発を進めています。

4. バイオプラスチック開発への適用事例

現在、私たちの生活には欠かせないプラスチックですが、CO₂排出量の削減や海洋プラスチックごみ問題への対策といった課題があります。NECでは、こうした課題にいち早く取り組み、2000年代の初めからバイオプラスチックの開発を行い、製品化の実績を重ねています(写真)³⁾。今後、環境問題の意識の高まりとともに更なる開



写真 NEC の作る新しいバイオプラスチック製品

発の加速が期待されています。

そこで、素材開発プラットフォームを適用して、実証を進めています。その実証実験として、バイオプラスチックの新素材の探索をする場合、ノーマルなマテリアルズ・インフォマティクス (MI) 技術を適用したときには、物性の予測が十分でなかったのに対して、第3章2節で紹介したコア技術を適用したことで、実用に足るレベルの物性予測の精度が得られるという結果を得ました。また、第3章1節でも述べた通り、バイオプラスチックに関する論文から LLM を通じて、原料名や強度、生分解性などから候補となる分子構造を示し、その製造方法を詳細に示すといった実例も得られています。このように、これらのコア技術を有効に用いることで、新たなバイオプラスチックを効率的に開発し、持続可能な社会づくりに貢献することを目指しています。

5. むすび

本稿では、素材開発分野への LLM 応用の取り組みとして、素材開発プラットフォームを紹介しました。今後、更なるお客様への迅速な対応が求められていく素材産業に対して、NEC はこれら2つのコア技術からなる AI を用いて素材開発を革新していきたいと考えています。

参考文献

- 1) Bhushan Kotnis, Kiril Gashteovski, Daniel Rubio, Ammar Shaker, Vanesa Rodriguez-Tembras, Makoto Takamoto, Mathias Niepert, Carolin Lawrence: MILIE: Modular & Iterative Multilingual Open Information Extraction, Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp.6939-6950, 2022
<https://aclanthology.org/2022.acl-long.478/>
- 2) Gorjan Radevski, Kiril Gashteovski, Chia-Chien Hung, Carolin Lawrence, Goran Glavas: Linking Surface Facts to Large-Scale Knowledge Graphs, CoRR 2023
<https://arxiv.org/abs/2310.14909>
- 3) NEC: NEC の作る、新しいバイオプラスチック NeCycle®
<https://www.necplatforms.co.jp/product/necycle/index.html>

執筆者プロフィール

小淵 喜一

セキュアシステムプラットフォーム
研究所
主任

船矢 幸一

NEC Laboratories Europe
Chief Technology Architect

當山 清彦

セキュアシステムプラットフォーム
研究所
リードリサーチエンジニア

田中 修吉

セキュアシステムプラットフォーム
研究所
ディレクター

ダニエル オノロルビオ

NEC Laboratories Europe
Senior Research Engineer

マーティン レンチェン ミン

NEC Laboratories America
Department Head

イアン メルビン

NEC Laboratories America
Researcher

関連 URL

研究開発型企業に求められる R&D の明日を徹底討論 先駆者たちに学ぶデータ活用の最前線

<https://events.nikkeibp.co.jp/event/2023/bpi230706/>

NEC 技報のご案内

NEC 技報の論文をご覧くださいありがとうございます。
ご興味がありましたら、関連する他の論文もご一読ください。

NEC技報WEBサイトはこちら

NEC技報（日本語）

NEC Technical Journal（英語）

Vol.75 No.2 ビジネスの常識を変える生成AI特集 ～社会実装に向けた取り組みと、それを支える生成AI技術～

ビジネスの常識を変える生成AI特集によせて
生成AI技術への取り組み ～基盤から応用、ルール作りまで～

◇ 特集論文

急速に広まる生成AIの市場適用

NECにおける生成AIの取り組みについて
電子カルテと医療文書の作成支援による医師業務効率化
映像×LLMによる報告書作成業務の自動化
映像分析と生成AIによるリアル世界の行動理解
サイバー脅威インテリジェンス生成自動化
生成AIの社内活用を進めるNEC Generative AI Service (NGS)
ソフトウェア・システム開発への生成AIの活用
LLMとMIで革新する素材開発プラットフォーム
LLMと画像分析を活用した被災状況の把握

生成AIの可能性を高める基盤技術

日本語性能に優れたNECの大規模言語モデル (LLM)
NECの生成AIを支える国内企業で最大規模のAIスーパーコンピュータ
より安全な大規模言語モデル (LLM) を目指して
データを秘匿したまま連携できる連合学習技術とLLMへの適用可能性
大規模言語モデル (LLM) によるFew-shotクラスタリングの可能性
オープンドメイン常識推論のための知識拡張型プロンプト学習
AI連携とオーケストレーションのための基盤ビジョンLLM
クエリを考慮した新規手法により関連する企業データを減らしLLM APIの使用コストを最適化

AI技術が社会へ浸透するために

AI標準化・ルールメイクに関する動向とNECの取り組み
人権尊重に向けたNECのAIガバナンスの取り組み
RCModelを用いたAIリスクマネジメントのための人材育成事例

◇ NEC Information

2023年度C&C賞表彰式典開催



Vol.75 No.2
(2024年3月)

特集TOP