

仕 様 書

1. 件名

有機分子データに対する深層学習モデルの開発補助

2. 研究の概要・目的

2-1. 概要・目的

国立研究開発法人産業技術総合研究所人工知能研究センター（以下「産総研」という。）では、国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）予算「AI セーフティ強化に関する研究開発」における研究テーマである「事前知識を活用した科学情報のための AI モデルのセーフティ管理技術に関する研究開発」を行っている。本作業では、有機分子データに対する深層学習モデルの開発補助として、すでに論文発表され一般に公開されている複数のモデルの調査と整備を行う。整備されたそれぞれの深層学習モデルおよびそのプログラムは、産総研が独自に開発するモデルとの精度比較に利用する。この比較結果は、論文の発表とソースコードの公開の際に必要となる。

2-2. 用語の定義

本仕様書で使用される用語とその意味について、以下に記す。

カテゴリ	用語	説明
組織及び人物	産総研担当者	本システムの企画及び運用等を担当する者及び所管部署の業務運用担当者。
	調達担当者	本調達の契約手続き等を担当するもの。
	受注者	本調達の対象となる業務に従事する事業者。
その他	情報セキュリティインシデント	産総研が望まない単独若しくは一連の情報セキュリティ事象、又は予期しない単独若しくは一連の情報セキュリティ事象であって、事業運営を危うくする確率及び情報セキュリティを脅かす確率が高いもの。
	情報セキュリティポリシー	産総研の情報セキュリティ基本方針、情報セキュリティ規程、情報セキュリティ実施要領及び情報セキュリティ実施ガイドの総称。

3. プログラムの概要

本件は産総研の研究開発に必要となる、既存の複数の有機分子データに対する深層学習モデルのプログラム群である。

4. システム開発の背景（経緯、属性）

これまで、科学データに対する人工知能の研究開発を目的として、有機分子データに対する機械学習モデルが開発されてきた。特に近年は、言語あるいは画像データの深層学習モデルの発展に伴い、有機分子データにおいても多種多様な大規模基盤モデルが開発・公開されるようになってきた。このように公開されている既存のモデルと、産総研が独自に開発しているモデルとの精度を効率的に比較・検証することで、モデル開発を迅速に進めることを目指す。

5. 開発内容構成

5-1. 複数の深層学習モデルの調査

5-2. プログラムへの変換・整備

6. 構成毎の開発仕様

6-1. 複数の深層学習モデルの調査

以下 3 つの深層学習モデルの調査を行い、進捗報告の際にスライド資料により報告を行うこと。

調査項目として、モデルがどのようなフォーマットの入力データを必要とするのか、モデルをどのような設定で学習やファインチューニングをするのか、そしてモデルの出力結果がどのようなフォーマットで得られるのか等をまとめること。これらの項目について、以下の 3 つの最新の深層学習モデルを調査すること。

- ・Graphormer (Microsoft 社が開発した分子のグラフトランスフォーマーモデル。既存のグラフニューラルネットワークを含めた、様々なモデルのベースライン。<https://github.com/microsoft/Graphormer>)

- ・Uni-Mol+ (AlphaFold をベースとした高精度な分子トランスフォーマー。ベンチマークでは現時点で最高精度。https://github.com/deepmodeling/Uni-Mol/tree/main/unimol_plus)

- ・Transformer-M (トランスフォーマーのみを用いたシンプルだが高精度なモデル。トランスフォーマーのベースライン。<https://github.com/ljsj2408/Transformer-M>)

6-2. プログラムへの変換・整備

(1) 上記の 3 つのモデルのプログラムを、産総研が指定する動作内容（モデ

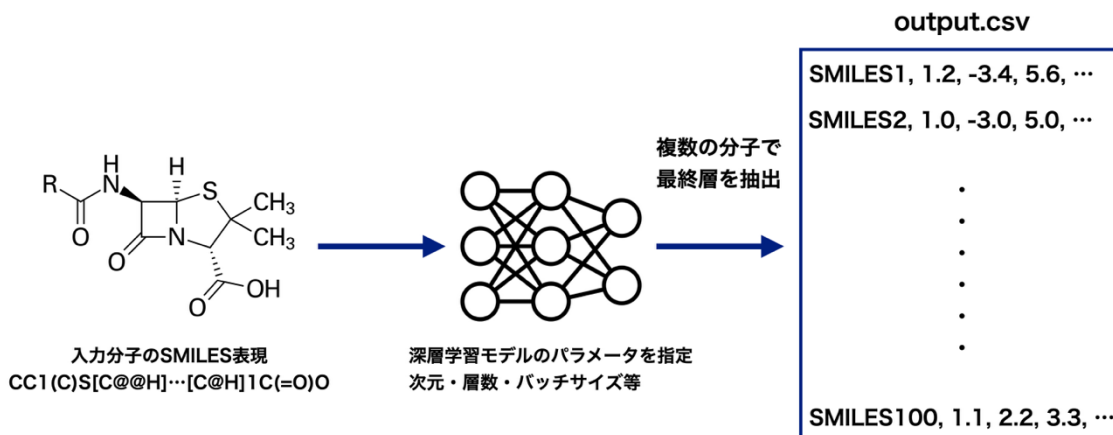


図 1：プログラムの入出力の概要図。複数の分子の SMILES データを入力し、モデルの出力層を抽出し、最終的に CSV ファイルとして出力すること。

ルの学習・ファインチューニングや出力結果の保存・再利用など）が実行できるプログラムに変換・整備すること。そのプログラムでは、設定したパラメータ（層数・次元・バッチサイズ等）の深層学習モデルについて、その出力層を csv ファイルとして出力すること。出力層は、一つの分子につき一つの多次元ベクトル、あるいは行列となるため、複数の分子をまとめた csv ファイルとして出力すること。プログラムの概要を図 1 に示す。

(2) 具体的な作業内容として、以下の 3 点の要件を満たすこと。

1) 入力データは分子データの SMILES (図 1 左) となるが、これ以外にも追加で必要な入力データがある場合 (3 次元構造等) は、SMILES からそれを作成した上で入力すること (図 2 左)。

2) 出力の最終層の抽出 (図 1 右) について、物性値を学習・予測する直前の層 (ベクトル) になるが、それ以外のデータが出力される場合 (3 次元構造等) は、それを調査し適切なフォーマットで保存・提供すること (図 2 右)。

3) 産総研が貸与するデータ (数百から千サンプル程度) でファインチューニングを行い、動作を確認すること。ファインチューニングとは、学習済みモデルのパラメータを初期値として、新たなデータでパラメータを学習することを意味する。

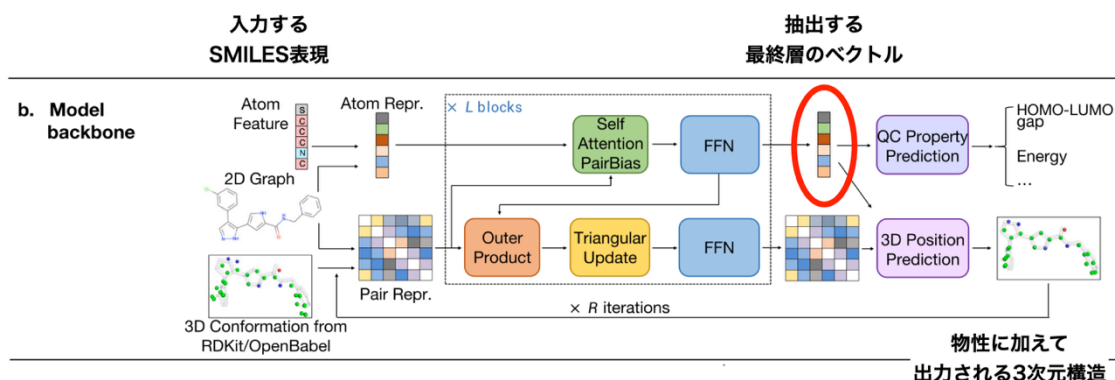


図 2 : Uni-Mol+の概略図。入力が分子の SMILES、出力がベクトルとなる。このベクトルから分子の物性値を学習・予測する。加えて、分子の 3 次元構造も出力される。

7. プログラム作成の条件等

7-1. プログラム作成使用言語及び動作環境等

- ① プログラミング言語や開発環境は、Python、PyTorch、RDKit を用いること。その他、デバッグや可視化などを行う場合は、適宜専用のライブラリを用いること。
- ② 2 週間に 1 度程度、整備したプログラムの説明とスライド資料による進捗報告（主にオンライン会議）を行うこと。

8. 発注側で貸与するデータ及び容量等

(1) プログラムの動作検証用の分子データセット

9. 特記事項

- (1) サプライチェーン・リスクに対応するため、別紙に記載する事項に従って契約を履行しなければならない。

10. 完成品の試験・確認

発注者は、スライド資料に記載されている操作手順を実際に行うなどして、プログラムの完成度と品質を確認する。

- ① プログラムの完成度と品質は、産総研が提供するデータセットで動作確認を行うことで確認する。加えて、進捗報告でのスライド資料により調査したモデルの詳細を確認する。
- ② 受注者は、上記確認作業において発注者を支援すること。

1 1. 納入の完了

本システムは、「1 2. 納入物品」に記載された納入物品が過不足なく納入され、仕様書を満たしていることを確認して、納品の完了とする。受注者は確認にかかる作業を支援すること。

1 2. 納入物品

- (1) 変換・整備プログラム 一式
- (2) 進捗報告スライド資料 1 部

* 共有ストレージからのダウンロードによる。

1 3. 納入期限及び納入場所

納入期限：令和 7 年 12 月 19 日

納入場所：東京都江東区青海 2-4-7

国立研究開発法人産業技術総合研究所

人工知能研究センター

臨海副都心研究センター別館 9 階 09202 室

1 4. 成果の取扱い

- (1) 国立研究開発法人産業技術総合研究所（以下「産総研」という。）は、受注者がプログラム作成により得られた技術上の成果のうち産総研が指示するもの（以下「成果」という。）についての利用及び処分に関する権利を専有するものとする。ただし、当該成果のうち乙の所有するプログラム、関連ドキュメント、ノウハウ、それらの関連資料の著作物の著作権は、乙に留保されるものとするが、甲が成果を利用するために必要な範囲で、甲に対して当該著作権の利用を許諾するものとする。
- (2) 受注者は、成果に係るソフトウェアの著作権（著作権法第 27 条及び第 28 条に規定する権利を含む。）及び意匠登録を受ける権利を産総研に譲渡するものとし、著作者人格権を行使しないものとする。ただし、パッケージ製品に係るものは除く。
- (3) 受注者は、産総研に対し、納品した成果品が第三者の知的財産権を侵害しないことを保証するものとする。なお、納品した成果品について、第三者の権利侵害の問題が生じ、その結果、産総研又は第三者に費用や損害が生じた場合は、受注者は、その責任と負担においてこれを処理するものとする。ただし、かかる侵害が産総研の指示によって発生した場合又は産総研の責に帰すべき事由による場合はこの限りではない。

15. セキュリティ要件

15.1. 情報セキュリティポリシーに関する要件

- ① 本業務の履行に当たっては、産総研の情報セキュリティポリシー（別途定める読み替え条項に従うものとする。以下同じ。）を遵守するとともに、情報セキュリティポリシーにおいて産総研に求められる水準の情報セキュリティ対策を講じること。なお、産総研の情報セキュリティ規程については、下記 URL を参照のこと。その他の情報セキュリティポリシーの詳細については受注者決定後に提示する。

【国立研究開発法人産業技術総合研究所情報セキュリティ規程】

https://www.aist.go.jp/Portals/0/resource_images/aist_j/outline/comp-legal/pdf/securitykitei.pdf

- ② 産総研の情報セキュリティポリシーの見直しが行われた場合は、見直しの内容に応じた情報セキュリティ対策を講じること。なお、対応内容については産総研担当者に事前に報告し承認を得ること。

15.2. その他セキュリティに関する要件

- ① 受注者は、本業務の履行に際して、秘密である旨を示されて提供を受けた秘密情報を秘密として適切に保持することとし、第三者に開示又は漏洩してはならない。
- ② 受注者は、本業務の履行によって知った一切の情報を本業務の履行以外の目的に利用してはならない。契約終了後も同様とする。
- ③ 提供する資料は産総研担当者の了解なしに所外に持ち出してはならない。
- ④ 産総研の所外へ持ち出した資料については一覧を作成し、産総研担当者に提出すること。なお、契約終了後、速やかに返却または廃棄し、産総研担当者に報告すること。
- ⑤ 受注者は、契約締結後、情報セキュリティ管理体制を記載したドキュメントを産総研担当者に提出すること。
- ⑥ 受注者は、本業務において、受注者の従業員若しくはその他の者によって、意図せざる変更が加えられない管理体制とすること。
- ⑦ 受注者は、産総研の求めに応じて、資本関係、役員等の情報、委託事業の実施場所並びに委託事業従事者の所属、専門性（情報セキュリティに係る資格・研修実績等）、実績及び国籍に関する情報提供を行うこと。
- ⑧ 本業務にかかる情報に関する情報セキュリティインシデントが生じた場合、速やかに報告の上、原因の分析を実施し、産総研担当者と対処内

容及び再発防止策を検討すること。当該インシデントへの対処を実施するにあたっては、事前に産総研担当者の確認を得ること。

- ⑨ 情報セキュリティインシデントが生じたことで、受注者の作業環境等の確認が必要となった場合には、産総研の調査に協力を行うこと。
- ⑩ 本業務の履行における情報セキュリティ対策の履行状況を確認するため、産総研が提示するチェックリストの内容に基づき、定期的に情報セキュリティ対策の履行状況を報告すること。
- ⑪ 産総研担当者より、情報セキュリティ対策の履行が不十分であると指摘された場合は、速やかに是正処置を講ずること。
- ⑫ 本業務の履行における情報セキュリティ対策の履行状況を確認するために、産総研が情報セキュリティ監査の実施を必要と判断した場合、受注者は、産総研が定めた実施内容（監査内容、対象範囲、実施者等）に基づく情報セキュリティ監査を受け入れること。
- ⑬ 受注者は、産総研の許可なく、本業務の一部又は全部を第三者（再委託先）に請け負わせてはならない。ただし、受注者に求めている情報セキュリティ対策を、再委託先が実施することを再委託先に担保させるとともに、再委託先の情報セキュリティ対策の実施状況を確認するために必要な情報を産総研に提供し、承認申請書を提出して、事前に産総研の書面による承認を受けた場合はこの限りではない。

16. 付帯事項

- ・ 受注者は、産総研担当者の求めにより、作業の進捗状況及び作業内容について報告（2週間1度程度）しなければならない。
- ・ 納入されたプログラム等における発注側の責めによらない1年以内の動作不良等不具合については、その補修、調整等責任をもって無償で速やかに行うこと。
- ・ 本仕様書の技術的内容に関する質問等については、産総研担当者と協議すること。
- ・ 本仕様書に定めのないこと項及び疑義が生じた場合は、調達担当者と協議のうえ決定する。

サプライチェーン・リスク対応に係る特記事項

1. サプライチェーン・リスクへの対応

受注者は、機器等の意図的な不正改造及び情報システム又はソフトウェアに不正なプログラムを埋め込むなど、国立研究開発法人産業技術総合研究所（以下、「産総研」という。）の意図しない変更が加えられたときに生じ得る情報の漏えい若しくは破壊又は機能の不正な停止、暴走その他の障害等の情報セキュリティ上のリスク（以下「サプライチェーン・リスク」という。）に対応するため、受注者は「IT 調達に係る国の物品等又は役務の調達方針及び調達手続に関する申合せ」（平成 30 年 12 月 10 日関係省庁申合せ）に基づく対応を図らねばならない。

2. 意図しない変更に対する対策

- ①受注者は、本業務の履行に際して、サプライチェーン・リスクが潜在すると知り、又は知り得るべきソースコード、プログラム等（以下「ソースコード等」という。）の埋込み又は組込みその他産総研担当者の意図しない変更を行ってはならない。
- ②受注者は、本業務の履行に際して、サプライチェーン・リスクが潜在すると知り、又は知り得るべきソースコード等の埋込み又は組込みその他産総研担当者の意図しない変更が行われないうに相応の注意をもって管理しなければならない。
- ③受注者は、本業務の履行に際して、情報の窃取等により研究所の業務を妨害しようとする第三者から不当な影響を受けるおそれのある者が開発、設計又は製作したソースコード等（受注者がその存在を認知し、かつ、サプライチェーン・リスクが潜在すると知り、又は知り得るべきものに限り、主要国において広く普遍的に受け入れられているものを除く。）を直接又は間接に導入し、又は組み込む場合には、これによってサプライチェーン・リスクを有意に増大しないことを調査、試験その他の任意の方法により確認又は判定するものとする。

3. サプライチェーン・リスクにかかる調査の受入れ体制

- ①受注者は、本業務に産総研担当者の意図しない変更が行われるなど不正が見つかったときは、追跡調査や立入検査等、産総研と連携して原因を調査し、サプライチェーン・リスクを排除するための手順及び体制を整備し、当該手順及び体制を示した書面を産総研担当者に提出しなければならない。

4. サプライチェーン・リスクを低減するための対策

- ①受注者は、サプライチェーン・リスクを低減する対策として、本業務の設計、構築、運用・保守の各工程における不正行為の有無について定期的または必要に応じて監査を行う体制を整備するとともに、本業務により産総研に納入する納入物品に対して意図しない変更が行われるリス

クを回避するための試験を行わなければならない。当該試験の項目は、情報セキュリティ技術の趨勢、対象の情報システムの特性等を踏まえ、受注者において適切に設定するものとする。

- ②機器の納入であり、かつ、設計、構築、運用・保守の各工程が存在しない場合は、4. ①の対応は不要。

5. 受注者の業務責任者等

- ①受注者は、本業務の履行に従事する業務責任者及び業務従事者（契約社員、派遣社員等の雇用形態を問わず、本業務の履行に従事する全ての従業員をいう。以下同じ。）を必要最低限の範囲に限るものとする。
- ②機器納入であり、かつ、設計、構築、運用・保守の各工程が存在しない場合は、5. ①の対応は不要。

6. 再委託

6.1 本業務の第三者への委託の制限

受注者は、産総研の許可なく、本業務の一部又は全部を第三者（再委託先）に請け負わせてはならない。ただし、6.2 に定める事項を遵守する場合はこの限りではない。

6.2 第三者への委託に係る要件

- ①受注者は、本業務の一部又は全部を第三者に再委託するときは、再委託先の事業者名、住所、再委託対象とする業務の範囲、再委託する必要性について記載した承認申請書を、委託元である産総研に提出し、書面による事前承認を受けなければならない。
- ②受注者は、本業務の一部又は全部を第三者に再委託するときは、再委託した業務に伴う再委託者の行為について、全ての責任を負わなければならない。
- ③受注者は、知的財産権、情報セキュリティ（機密保持を含む。）及びガバナンス等に関して、本仕様書が定める受注者の責務を再委託先も負うよう、必要な処置を実施し、その内容について委託元である産総研の承認を得なければならない。
- ④受注者は、受注者がこの仕様書の定めを遵守するために必要な事項について本仕様書を準用して、再委託者と約定しなければならない。
- ⑤受注者は、前号に掲げる情報の提供に加えて、再委託先において本委託事業に関わる要員の所属、専門性（情報セキュリティに係る資格・研修実績等）、実績及び国籍についての情報を委託元である産総研へ提出すること。
- ⑥受注者は、再委託先において、産総研の意図しない変更が加えられないための管理体制について委託元である産総研に報告し、許可又は確認（立入調査）を得ること。

7. その他

- ①提出された資料等により産総研担当者に報告された内容について、サプライチェーン・リスクが懸念され、これを低減するための措置を講じる必要があると認められる場合に、調達担当者は

受注者に是正を求めることがあり、受注者は相当の理由があると認められるときを除きこれに応じなければならない。

- ②産総研は、受注者の責めに帰すべき事由により、本情報システムに産総研担当者の意図しない変更が行われるなど不正が見つかった場合は、契約条項に定める契約の解除及び違約金の規定を適用し、本業務契約の全部又は一部を解除することができる。