

# 事前学習言語モデルとグラフニューラルネットワークの組合せによる専門知識の抽出

浅野 聖也<sup>1</sup> 磯沼 大<sup>1,2</sup> 浅谷 公威<sup>1</sup> 野村 美鈴<sup>3</sup> 森 純一郎<sup>1,4</sup> 坂田 一郎<sup>1</sup>

<sup>1</sup> 東京大学 <sup>2</sup> エディンバラ大学 <sup>3</sup> ダイキン工業株式会社 <sup>4</sup> 理研 AIP

seiyaasano4084@u-tokyo.ac.jp {isonuma, isakata}@ipr-ctr.t.u-tokyo.ac.jp

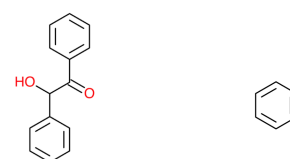
asatani@tmi.t.u-tokyo.ac.jp misuzu.harada@daikin.co.jp mori@mi.u-tokyo.ac.jp

## 概要

本研究では、事前学習言語モデルとグラフニューラルネットワークを組み合わせることで、化学物質に関する知識抽出を試みる。具体的には、マスク言語モデルの一つである BERT にグラフニューラルネットワークを組み込むことで、化学物質の分子構造をグラフ情報として活用することを可能にし、汎化性能が向上することを示す。実験では Scopus の科学論文抄録コーパスを利用し、特に学習コーパスに出現する頻度が少ない化学物質について、従来のモデルに比べて顕著な性能向上を確認した。

## 1 はじめに

近年の自然言語処理分野では、学習済み言語モデルの活用による飛躍的な進歩が見られる。中でも、BERT (bidirectional encoder representations from transformers) [1] は、多様なタスクで優れた成果を上げており、その柔軟性と拡張性により広範な応用が提案されている。その一つとして、学習済み BERT を知識抽出に用いる方法が注目を集めている [2]。これは BERT が文章中のマスクされたトークンを予測するように学習されていることを利用したものである。“[MASK] is the president of the United States of America in 2023.” のように、あるエンティティに関する文章についてエンティティ部分をマスクし、学習済み BERT にマスクされた部分に入るトークンを予測させることで、自然言語をクエリとした柔軟な知識抽出を行うことができる。従来の知識抽出手法では、エンティティの性質やエンティティ同士の関係を考慮した知識グラフを人手で構築する必要があるため、学習済み言語モデルを知識抽出に活用できれば、既存手法に比べて低コストかつ柔軟な知識抽出が可能になる。



名称	Benzoin	Benzene
分子量	212.24 g/mol	78.11 g/mol
沸点	176.2 °F at 760 mmHg	651 °F at 760 mmHg
溶解度	less than 0.1 mg/mL at 64 °F	1 to 5 mg/mL at 64 °F

図 1 名称は類似するが、異なる物性を示す物質例

ただし、このアプローチは専門的なドメインでは有効でないことが指摘されている [2, 3]。特に化学分野では、化学物質の解釈に専門的な知識や複雑な構造解析を伴う場合が多く、言語モデルでそれらを扱うことは難しい。例えば、図 1 に示す二つの化学物質はいずれも名称は類似しているが、全く異なる物性を示す。<sup>1)2)</sup> こうした化学物質を言語モデルで扱うことは困難であり、化学物質に関する専門知識や構造情報を活用する必要がある。これまでの研究では、化学物質の構造情報を用いて物性を学習・予測する試み [4, 5] や、化学物質の分子構造を曖昧性なく文字列化した記法である SMILES 表現を言語モデルで扱う手法 [6] などに取り組まれているが、化学物質の分子構造を言語モデルによる知識抽出タスクに利用した例は確認されていない。

本研究では、化学物質に関する知識や分子構造を扱うため、化学物質の構造情報を BERT に入力する手法を提案する。具体的には、化学物質のグラフ構造情報をグラフニューラルネットワークの一つである Graphormer [7] に入力し、その出力を BERT の入力として用いる。専門的なドメインのコーパスを対象に知識抽出タスクを解かせることで、Graphormer と

1) <https://pubchem.ncbi.nlm.nih.gov/compound/8400>

2) <https://pubchem.ncbi.nlm.nih.gov/compound/241>

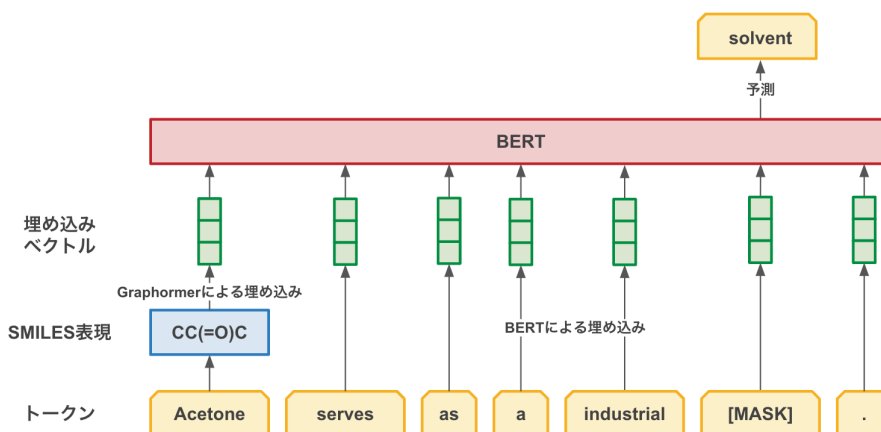


図2 提案手法の概要

BERT を一体的に学習する。これにより、化学物質の専門知識や構造情報を扱うことを可能にし、化学分野での知識抽出手法として BERT の活用を目指す。

評価実験では、PubChem から取得した化学物質の説明文章のマスク部を予測する知識抽出タスクにて、Graphormer を用いたことによる効果を評価した。その結果、特に学習コーパスに出現する頻度が少ない化学物質について、Graphormer を用いていないモデルに比べて顕著な性能向上を確認した。

## 2 提案手法

本節では、Graphormer を組み合わせることで、BERT を化学分野の知識抽出タスクに適用する方法について述べる。図2 に提案手法の概要を示す。

### 2.1 文章中の化学物質の特定

まず、入力文章中に含まれる化学物質を特定する方法について説明する。本手法では ChemDataExtractor[8] を用いて入力文章中のトークンのうち化学物質のものを特定する。さらに、National Center for Biotechnology Information (NCBI) の PubChem データベースを用いて、抽出した化学物質を SMILES 表現に変換する。

ただし、Graphormer によって埋め込みが作成できない単原子分子や、異常原子価の物質は抽出しない。

### 2.2 化学物質の埋め込み

変換された SMILES 表現を、Graphormer により埋め込みベクトルに変換する。Graphormer は、グラフ構造データを処理するために設計された transformer ベースのモデルであり、特に分子構造などの複雑なグラフ表現の解析に適している。また、Graphormer

は、大規模な分子データセット上での事前学習を通じて、幅広い化学物質に対する汎化能力を持つ。これにより、学習コーパス中の出現回数が少ない化学物質でも、知識抽出を行えることが期待される。

まず、与えられた SMILES 表現をもとに、原子をノード、その結合をエッジとしたグラフ情報を作成する。グラフ情報には、各ノードに当たる原子について、原子番号やキラリティを含む9つの情報が用いられ、エッジにあたる結合部分については、単結合や二重結合といった結合の種別に関する情報が含まれる。そして、ノード  $i, j$  について以下の式 (1), (2) に示すようにアテンションを計算し、式 (3) の様に層ごとに各ノードの特徴ベクトルを計算する。

$$\mathbf{h}_i^{(0)} = \mathbf{x}_i + \mathbf{z}_{v_i}^{\text{deg-}} + \mathbf{z}_{v_i}^{\text{deg+}} \quad (1)$$

$$A_{ij} = \frac{(\mathbf{h}_i W^Q)(\mathbf{h}_j W^K)^T}{\sqrt{d_k}} + b_{\phi(v_i, v_j)} + c_{ij} \quad (2)$$

$$\mathbf{h}_i^{(l+1)} = \text{FFN} \left( \text{MHA} \left( \text{LN}(\mathbf{h}_i^{(l)}) \right) + \mathbf{h}_i^{(l)} \right) \quad (3)$$

- $\mathbf{h}_i^{(l)}$ : ノード  $i$  の  $l$  層目における特徴ベクトル
- $\mathbf{x}_i$ : ノード  $i$  の特徴ベクトル
- $\mathbf{z}_{v_i}^{\text{deg-}}$  と  $\mathbf{z}_{v_i}^{\text{deg+}}$ : ノード  $v_i$  の入次数と出次数に対応する埋め込みベクトル
- $A_{ij}$ : ノード  $i$  からノード  $j$  へのアテンション
- $W^Q$  と  $W^K$ : クエリとキーの重み行列
- $d_k$ : キーの次元数
- $b_{\phi(v_i, v_j)}$ : ノード  $v_i$  と  $v_j$  間の最短経路距離に基づくバイアス項
- $c_{ij}$ : ノードペア  $(v_i, v_j)$  に対する edge encoding 項
- FFN: フィードフォワードネットワーク

- MHA: マルチヘッドアテンション
- LN: レイヤー正規化

ここで注目すべき点は、ノード情報としてエッジの入出次数を考慮することで attention 機構においてノード重要性を捉える試み (Centrality Encoding), ノード間の最短経路の距離に基づいたバイアス項を追加しグラフの構造情報を捉える試み (Spatial Encoding), そして最短経路上の各エッジについてその特徴ベクトルと重みベクトルの内積を平均したものを追加しノード間の関係性を捉える試み (Edge Encoding) である。このように計算された全ノードについての特徴ベクトルを READOUT 層に通し、グラフ全体の埋め込みを作成している。Graphormer により出力されるベクトルは、BERT のトークンの埋め込みベクトルと同様に 768 次元のベクトルであるため、BERT の学習においてトークンの埋め込みベクトルと同様に扱うことができる。

## 2.3 Graphormer と BERT の学習

Graphormer により出力された化学物質の埋め込みベクトルと各トークンの埋め込みベクトルをもとに、BERT はマスクされたトークンの確率を推定する。そして、マスクされたトークンの尤度を最大化するように BERT と Graphormer のパラメータを更新する。これにより、Graphormer は学習時に出現しなかった化学物質に対しても、既に学習した化学物質の情報からその物性を推定することができ、BERT はその推定された物性に関する情報をもとに、知識抽出タスクを解くことができると期待される。

## 3 実験

### 3.1 学習用データセット

学習用データセットとして、Scopus データベースから取得した科学論文の抄録コーパスを用いた。Scopus はその広範な論文収蔵量で知られ、多様な科学分野にわたる豊富な文章データを提供している。本研究では Scopus に収蔵されている 2022 年までの論文抄録データ 63,584,937 件のうち、2.1 節で説明した方法で埋め込み可能な化学物質が含まれる抄録 19,037,080 件を使用した。

### 3.2 モデル

学習を行うベースモデルとして、SciBERT [9] を用いた。SciBERT は学術文献に特化して事前学習され

た BERT モデルであり、一般的な BERT モデルと比較して学術文献の処理において優れた性能を示す。このモデルは、学術文献で使用される専門用語と文体に適応するために、EBM-NLP [10] や SciERC [11] といった学術論文からなる大規模なコーパスを使用して事前学習されている。従って、本研究で扱う化学物質名の文脈的意味解析や、科学的テキストの精密な情報抽出に適していると考えられる。

一方、Graphormer は PCQM4Mv2 にて事前学習されたものを用いた。PCQM4Mv2 は、PubChemQC プロジェクト [12] から派生した、分子グラフに関する量子化学的特性の予測のための大規模なデータセットである。約 380 万の分子に対してラベル付けされた、HOMO-LUMO ギャップの予測タスクで事前学習された。HOMO-LUMO ギャップは分子の電子的特性と化学反応性に影響するため [13]、このタスクを解くことで、Graphormer が分子構造と分子の化学的性質との関係を学習することが期待される。

### 3.3 評価方法

本研究では評価用データセットとして、PubChem の PUG REST API を通じて得られた、学習時に登場した化学物質の説明文章 53,353 件を用いた。化学物質の説明文章は “XX is YY” という形式で提供されており、XX には対象となる化学物質名が、YY にはその説明が記載されている。これらのテキストは、モデルが化学物質に関する専門的な情報をどの程度正確に理解し処理できるかを評価するために用いられる。本研究では、既存研究 [2] と同様に [MASK] トークンに関する知識抽出タスクで性能を確認した。具体的には、YY 部分に含まれるトークンの中からランダムにマスクし、元のトークンを予測する。ただし、YY 部分は知識抽出対象の化学物質についての知識を表しているため、YY 部分をランダムにマスクすることで知識関係の理解を確認できると考えられる。本研究では、既存研究 [2] と同様に、評価指標として Precision@k を採用した。具体的には、モデルによる上位 k 番目までの予測が正解トークンを含んでいれば 1、そうでなければ 0 を出力し、全評価用データセットでの平均を求める。

### 3.4 比較モデル

本研究では、以下のモデルについて比較を行った。

1. 追加学習を行っていない SciBERT
2. Graphormer を使わず追加学習した SciBERT

表1 PubChem 上の化学物質説明文におけるトークン予測の Precision@k 比較 (学習時の登場回数ごとに評価)

学習時の登場回数	SciBERT w/o finetune		SciBERT w/o Graphormer		SciBERT w/ Graphormer	
	Precision@1	Precision@5	Precision@1	Precision@5	Precision@1	Precision@5
1-9	0.799	0.912	0.803	0.929	<b>0.811</b>	<b>0.944</b>
10-99	0.791	0.918	<b>0.802</b>	0.934	<b>0.802</b>	<b>0.947</b>
100-999	0.794	0.912	0.802	0.925	<b>0.804</b>	<b>0.937</b>
1000-9999	0.783	0.913	<b>0.797</b>	0.931	0.795	<b>0.932</b>
10000-	0.762	0.906	<b>0.772</b>	0.930	<b>0.772</b>	<b>0.933</b>
All	0.786	0.925	0.794	0.932	<b>0.801</b>	<b>0.945</b>

### 3. Graphormer を用いて追加学習した SciBERT

これらを比較することで, Graphormer を用いた追加学習が, どのようにモデルの性能に影響を与えるか明らかにする. 特に, 学習データ中の登場頻度が低い化学物質に対して, Graphormer の利用がモデルの汎化性能向上に寄与するか検証する.

## 4 結果

本実験では, 学習データセットでの化学物質の登場頻度に応じて, 異なる条件下でのモデルの性能評価を行った. 学習データ中の登場回数ごとに化学物質を分類し (1-9 回, 10-99 回, 100-999 回, 1000-9999 回, 10000 回以上), それぞれに対して分割した評価用データセットについて, Precision@1 と Precision@5 を計算し比較した. さらに, 評価用データセット全体を対象にした場合の性能評価も行った.

### 4.1 Precision@k

3.4 で述べた条件に基づき, Precision@k によるトークン予測精度比較の結果を表 1 に示す.

Graphormer を用いて追加学習した SciBERT が, ほぼ全ての登場回数において最も高い精度を出した. また, Graphormer を使わず追加学習した SciBERT についても, Precision@1 では 10-99 回, 1000-9999 回, 10000 回以上出現した化学物質について最大精度を出した. また, いずれの追加学習したモデルについても, 追加学習を行っていない SciBERT よりも高い精度を出した.

### 4.2 考察

表 1 に示した結果について注目すべき点は, 化学物質の登場回数が少ない場合 (1-9 回) における Precision@k の値である. このカテゴリにおいて, Graphormer を用いて追加学習した SciBERT は, 他の 2 つのモデルに比べ Precision@1, Precision@5 と

に高い精度を示した. これは, Graphormer を活用することで, 学習データでの登場頻度が低い場合においても化学物質の構造情報をより効果的に統合し, 理解を深めることができることを示唆している.

一方で, 化学物質が頻繁に登場するカテゴリ (1,000 回以上) では, 追加学習における Graphormer の有無に大きな差は見られず, 一部 Graphormer を用いない場合の精度が用いる場合よりも高くなっていた. これは, データセットに頻繁に登場する化学物質に関しては, 既存の SciBERT モデルでも十分な学習が可能であることを示している.

これらの結果は, データセット内で稀にしか登場しない化学物質を含む文章を扱う上で, Graphormer を用いた追加学習の有効性を強く支持している.

## 5 おわりに

本研究では SciBERT と Graphormer を例としてマスクド言語モデルとグラフニューラルネットワークを組み合わせることで, 化学物質に関する知識抽出の性能向上を確認した. これにより, 専門的なテキストにおける学習済み言語モデルを用いた知識抽出について, 専門的な用語を別途それに特化したモデルで埋め込むことの有効性が示唆された.

今後の展望として, 本手法の学習過程を GPT[14] のように, BERT をベースとしていない言語モデルでも採用した場合に有効であるかの調査の実施が考えられる. また, 他の科学分野への応用も検討されるべきである. 特に, 生物学や物理学など他の複雑な専門領域におけるテキスト解析に本手法を応用することで, 新たな知識の発見や理解の深化が期待される.

## 謝辞

本研究は、ダイキン工業株式会社、NEDO JPNP20006 及び JST CREST JPMJCR21D1 の支援を受けたものである。

## 参考文献

- [1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [2] Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander Miller. Language models as knowledge bases? In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, pp. 2463–2473, Hong Kong, China, November 2019. Association for Computational Linguistics.
- [3] Mujeen Sung, Jinhyuk Lee, Sean Yi, Minji Jeon, Sungdong Kim, and Jaewoo Kang. Can language models be biomedical knowledge bases? In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 4723–4734, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.
- [4] Yu Rong, Yatao Bian, Tingyang Xu, Weiyang Xie, Ying Wei, Wenbing Huang, and Junzhou Huang. Self-supervised graph transformer on large-scale molecular data. **Advances in Neural Information Processing Systems**, Vol. 33, pp. 12559–12571, 2020.
- [5] Fang Wu, Dragomir Radev, and Stan Z Li. Molformer: Motif-based transformer on 3d heterogeneous molecular graphs. In **Proceedings of the AAAI Conference on Artificial Intelligence**, Vol. 37, pp. 5312–5320, 2023.
- [6] Gayane Chilingaryan, Hovhannes Tamoyan, Ani Tevosyan, Nelly Babayan, Lusine Khondkaryan, Karen Hambarzumyan, Zaven Navoyan, Hrant Khachatrian, and Armen Aghajanyan. Barts miles: Generative masked language models for molecular representations. **arXiv preprint arXiv:2211.16349**, 2022.
- [7] Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. Do transformers really perform badly for graph representation? **Advances in Neural Information Processing Systems**, Vol. 34, pp. 28877–28888, 2021.
- [8] Matthew C Swain and Jacqueline M Cole. Chemdataextractor: a toolkit for automated extraction of chemical information from the scientific literature. **Journal of chemical information and modeling**, Vol. 56, No. 10, pp. 1894–1904, 2016.
- [9] Iz Beltagy, Kyle Lo, and Arman Cohan. SciBERT: A pretrained language model for scientific text. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, pp. 3615–3620, Hong Kong, China, November 2019. Association for Computational Linguistics.
- [10] Benjamin Nye, Junyi Jessy Li, Roma Patel, Yinfei Yang, Iain Marshall, Ani Nenkova, and Byron Wallace. A corpus with multi-level annotations of patients, interventions and outcomes to support language processing for medical literature. In Iryna Gurevych and Yusuke Miyao, editors, **Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 197–207, Melbourne, Australia, July 2018. Association for Computational Linguistics.
- [11] Yi Luan, Luheng He, Mari Ostendorf, and Hannaneh Hajishirzi. Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii, editors, **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing**, pp. 3219–3232, Brussels, Belgium, October–November 2018. Association for Computational Linguistics.
- [12] Maho Nakata and Tomomi Shimazaki. Pubchemqc project: a large-scale first-principles electronic structure database for data-driven chemistry. **Journal of chemical information and modeling**, Vol. 57, No. 6, pp. 1300–1308, 2017.
- [13] Kenichi Fukui, Teijiro Yonezawa, and Haruo Shingu. A molecular orbital theory of reactivity in aromatic hydrocarbons. **The Journal of Chemical Physics**, Vol. 20, No. 4, pp. 722–725, 1952.
- [14] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. Improving language understanding by generative pre-training. 2018.

## A 知識抽出タスクの具体例

評価用データセットから取得した文章に対する予測結果例を掲載する。Dypyridamole についての説明文章

“2,3-dihydroxy-3-methylbutanoate is a hydroxy monocarboxylic acid anion. It is functionally related to a butyrate. It is a conjugate [MASK] of a 2,3-dihydroxy-3-methylbutanoic acid.”

の場合、正解単語 “base” に対する予測は以下のようになった。ただし、左から第1予測 - 第5予測である。

- BERT w/o finetune :  
[‘metabolite’, ‘form’, ‘product’, ‘compound’, ‘derivative’]
- BERT w/o Graphormer :  
[‘acid’, ‘ester’, ‘form’, ‘metabolite’, ‘product’]
- BERT w/ Graphormer :  
[‘acid’, ‘ester’, ‘form’, ‘metabolite’, ‘base’]

## B 学習用データセット前処理詳細

本研究での実験で用いたモデルの学習用データセットについては、Scopus の論文抄録コーパスを用いたが、そのうち埋め込み可能な化学物質を含む文章を選定するには以下のような流れで行った。

1. NIH(National Institutes of Health) の公開データ<sup>3)</sup> を活用して物質名と SMILES 情報を紐付けた辞書を作成した。それを利用して文章単語内に SMILES 変換できる化学物質があるもののみを抽出した。この操作により 63,584,937 件のデータが 50,981,559 件に剪定された。
2. ChemDataExtractor の Document 関数を利用して単語の化学物質判定を行った。化学物質判定される単語を抽出した後に、新たな化学物質 set にした。
3. 2で作成した set 内の化学物質を含む文章へさらに絞り込みを行った。この操作により 50,981,559 件のデータが 19,440,221 件に剪定された。
4. 2で作成したもののうち、graphormer を用いてベクトル化できない物質を除いて新たな set を作成した。
5. 4で作成した set 内の化学物質を含む文章へさらに絞り込みを行った。この操作により 19,440,221

3) <https://ftp.ncbi.nlm.nih.gov/pubchem/Compound/Extras/>

件のデータが 19,037,080 件に剪定された。

また、このデータセットを提案手法を適用したモデルに渡すため、文章中の 4 内に含まれる化学物質名部分を CHEM トークン ([CHEM]) というスペシャルトークンに置き換えた。スペシャルトークンにした意図は、学習時にマスクがされないようにするためである。そして、CHEM トークンと対応する置き換える前の化学物質名を index の形で保持し、モデルに渡すことで化学物質ごとの区別を行っている。

## C モデル実装詳細

モデルの実装には PyTorch<sup>4)</sup> と Hugging Face<sup>5)</sup> を用いた。事前学習済みモデルとして、SciBERT<sup>6)</sup>、Graphormer<sup>7)</sup> を用いた。ハイパーパラメータは学習を通して事前学習済みのモデルの設定を継承している。Graphormer の埋め込みを BERT の学習内で利用するため、Hugging Face の modeling\_bert.py<sup>8)</sup> の BertForLMaskedLM を継承し、入力トークンが 4 内に含まれる化学物質である場合は Graphormer による埋め込みを返すようにしている。さらに、Hugging Face の data\_collator.py<sup>9)</sup> の DataCollatorForLanguageModeling を継承し、モデルへの入力 batch に化学物質名と対応する index を渡すための処理を追加している。また、MLM での追加学習のコードは Hugging Face の run\_mlm.py<sup>10)</sup> をベースにした。当コードに対して、CHEM トークンをスペシャルトークンとして追加する処理を加えている。ただし、学習の際には、学習用データセットのうち、5% をバリデーションのために使用した。なお計算環境の都合上、評価用データセットについて、各登場回数及び全体について 1,000 件ランダムサンプリングして用いている。その際には seed 値を 0 に固定して行った。

4) <https://pytorch.org/>

5) <https://huggingface.co/>

6) [https://huggingface.co/allenai/scibert\\_scivocab\\_uncased](https://huggingface.co/allenai/scibert_scivocab_uncased)

7) <https://huggingface.co/clefourrier/graphormer-base-pcq4m4mv2>

8) [https://github.com/huggingface/transformers/blob/main/src/transformers/models/bert/modeling\\_bert.py](https://github.com/huggingface/transformers/blob/main/src/transformers/models/bert/modeling_bert.py)

9) [https://github.com/huggingface/transformers/blob/main/src/transformers/data/data\\_collator.py](https://github.com/huggingface/transformers/blob/main/src/transformers/data/data_collator.py)

10) [https://github.com/huggingface/transformers/blob/main/examples/pytorch/language-modeling/run\\_mlm.py](https://github.com/huggingface/transformers/blob/main/examples/pytorch/language-modeling/run_mlm.py)