# 特許請求の範囲の自動書き換え生成モデルのための大規模データセットの構築

河野 誠也 [1,2], 野中 尋史 [3], 吉野 幸一郎 [1,2]
[1] 理化学研究所ガーディアンロボットプロジェクト
[2] 奈良先端科学技術大学院大学 [3] 愛知工業大学
{seiya.kawano,koichiro.yoshino}@riken.jp, hnonaka@aitech.ac.jp

## 概要

本研究では，特許文書における特許請求の範囲の書き換えを自動生成することを目的とした書き換え生成モデルの基礎的検討を実施する。本研究では，このようなモデルを学習・評価するためのベンチマークとして，特定の特許出願に紐づけられた公開特許公報，特許公報からの情報の差分を取り込むことで，特許請求の範囲の書き換え事例を大量に収録したデータセットを構築した。次に，構築したデータセットを用いて大規模言語モデルに基づいた特許請求の範囲の自動書き換え生成モデルを構築し，その性能と限界について議論した。

## 1 はじめに

特許請求の範囲とは，特許を受けようとする発明を特定するための特許文書中の項目であり，一つ又は複数の「請求項」から構成される。請求項は，当該の発明が特許査定を受けるためや，既存の特許と比較してより強い権利を得るために，その記載内容を効果的に推敲することが重要である。しかなしながら，一般に，請求項の推敲は多くの時間と労力を要する作業となることが多いため，効果的な請求項の推敲支援ツールが求められている。これまでに，特許検索やパテントマップ生成などの基本的な特許情報処理のタスクに関する研究 [1, 2] や，特許請求の範囲の読解支援に関する研究 [3] は実施されている一方で，与えた特許請求の範囲を様々な観点からより良く書き換えるような書き換え生成モデルに関する実質的な研究は不十分である [4]。また，そのような書き換え生成モデルを学習・評価するためのデータセットも存在しない。

そこで，本研究では，特許請求の範囲（一つ以上の請求項で構成）の書き換えを自動生成することを目的とした書き換え生成モデルの基礎的検討を実施する。具体的には，与えられた特許請求の範囲から，より良い特許請求の範囲の書き換えを生成する書き換え生成モデルを大規模言語モデルに基づいて構築する。本研究では，このようなモデルを学習・評価するためのベンチマークとして，特定の特許出願に紐づけられた公開特許公報，特許公報からの情報の差分を取り込むことで，特許請求項の書き換え事例を大量に収録したデータセットを構築した。次に，構築したデータセットを用いて大規模言語モデルに基づいた特許請求の範囲の自動書き換えモデルを構築し，その性能と限界について議論する。

## 2 書き換えデータセット

特許請求の範囲の自動書き換えを目的とした書き換え生成モデルを学習・評価するためには，特許請求の範囲の書き換え事例（書き換え前の特許請求の範囲と書き換え後の特許請求の範囲のペア）を教師データとして大量に収録したデータセットが必要である。本研究では，このようなデータセットを構築するための書き換えの事例として，実際の特許出願に紐づけられた公開特許公報（特許登録の有無に関わらず一定期間後に公開される出願時の特許文書），特許公報（登録査定された特許文書）の特許請求の範囲のペアに着目した。一般に，多くの特許出願は一度拒絶された後に必要な修正（補正）を適用したのち登録査定される（図 1）。また，拒絶の有無に関わらず，手続き上必要な軽微な軽微な修正も実施される。特許制度の性質上，特許公報における特許請求の範囲は，特許を受けるために出願時の特許文書（特許公開公報がこれに対応）から必要な修正が適用されたものである。したがって，特許制度の観点においては，特許公報の特許請求の範囲は特許公開公報のものと比較して「より良い書き換え」の結果

として利用できる。

## 2.1 データセットの構築

本研究では，特許庁が提供する特許情報バルクデータ（過去分）をデータセット構築に用いた。具体的には，2004 年から 2022 年に特許庁のシステムを通じて公開された特許公開公報情報（A），特許公報情報（B9）の集合から，同一の特許出願番号を持つ公開公報（A），特許公報（B9）の特許請求の範囲のペアを抽出した。ここで，特許庁が配布する公報データは実際には xml 形式で提供されており，生成モデルの学習には適さない。したがって，公報 xml ファイルに XSLT 変換を適用し，J-PlatPat[1] 上での HTML view とほぼ等価なプレインテキスト形式に事前に変換した。さらに，特許出願番号に基づいて，特許庁が提供する特許情報標準データ（審査経過情報等），拒絶理由条文コード情報（特許審査官によって特許が拒絶される際にその根拠として引用された条文のコード情報），IIP パテントデータベース [5] と紐づけすることで，特許情報処理研究に利用可能な統合的なデータベースを整備した。実際には，特許請求の範囲だけではなく，発明の詳細内容の書き換えペアもデータセットに収録をしており今後の研究に利用可能である。

## 2.2 データセットの分析

データセットに収録されている特許公開公報（A）と特許公報（B9）の特許請求の範囲の有効なペア（両者に差分がある場合，ない場合），特許許公開公報（A）のうち未登録（未審査，拒絶，審査中など）であり対応する特許公報が存在しないものの内訳を表 1 に示す。ここで，XSLT 変換のエラーや元公報ファイルが破損しているような事例は除外していることに注意されたい。平均トークン数[2]，平均請求項数は，特許公報（B9）の特許請求の範囲のものであり，括弧は書き換え前の特許公開公報（A）からの異なり率を示している。未登録の場合は，公開公報（A）の統計を示している。

表 1 から，データセットには，合計で 4,856,533 件の特許出願に関する特許公開公報と特許公報のテキストペアが収録されている。ここで，特許公開公報（A）と特許公報（B9）のペア 2,245,640 件のうち，特許請求の範囲のテキストに差分がある場合が 1,839,880 件（81.9%），両者に差分がない場合が 405,760 件（18.1%）存在した。ここで，未登録の特許と比較して，登録された特許の方が平均トークン数と平均請求項数が減少する傾向が確認できた。請求項数は 21.7%程度（2 件程度）が減少している一方で，特許請求の範囲のテキストのトークン数は 5.8%程度しか減少していないことが確認できる。

**表 1** 構築したデータセットの内訳

| 公報種別 | 件数 | 平均トークン数 | 平均請求項数 |
|---|---|---|---|
| A [未登録] | 2,610,893 | 1034.41 | 7.84 |
| A, B9 ペア [差分なし] | 405,760 | 644.21 (+0.0%) | 6.03 (+0.0%) |
| A, B9 ペア [差分あり] | 1,839,880 | 739.83 (-5.8%) | 6.67 (-21.7%) |

また，A, B9 ペア [差分あり] の集合のうち，出願審査中に拒絶理由が発送された履歴があるものを対象とした平均編集距離（レーベンシュタイン距離）を表 3（付録を参照）に示す。ここで，line-dist, word-dist は編集距離を求める際の編集操作の単位として行単位，単語単位を用いた場合に相当する。。結果は，拒絶理由毎に編集距離の大きさが異なる傾向が確認できる。例えば，新規事項を追加する補正（第 17 条の 2 第 3 項），発明の単一性（第 37 条）のような拒絶理由において，特許公開公報と特許公報の間には大きな編集距離が存在している。一方で言い換えれば，出願人は出願した特許が特許査定を受けるために大幅な書き換えを適用したとみなすことができる[3]。これらの結果は，特許請求の範囲の書き換えの目的に応じた書き換えを実現することの重要性を示唆している。

# 3 書き換え生成モデル

本節では，特許請求の範囲の書き換え生成タスクを定義し，大規模言語モデルに基づいた特許請求の範囲の書き換え生成モデルの学習方法について述べる。

## 3.1 タスク設定

本研究が扱う特許請求の範囲の書き換え生成タスクは書き換え前の特許請求の範囲 $c$ のテキストを入力として与えたときに，与えた特許請求の範囲をより良く書き換えた特許請求の範囲 $c'$ のテキストを出力として生成することである。このような書き換え生成モデルを学習するためには，以下の目的関数を最小化する必要がある。

1） https://www.j-platpat.inpit.go.jp/
2） open-calm モデル付属の tokenizer によるサブワード分割結果を用いた
3） 実際の書き換え事例は付録を参照

$$L = -\sum_{(c,c')\in C} \log P(c'|c;\theta) \quad (1)$$

ここで，$C$ は書き換え前の特許請求の範囲 c と書き換え後の特許請求の範囲 $c'$ のペアの集合（学習データ），$\theta$ は書き換え生成モデルの学習可能なパラメタ集合である。さらに，想定される特許の拒絶理由などを，書き換え前の特許請求の範囲 c の付加情報として与えてもよい。3 節でも述べたように，このような書き換え生成モデルを学習するためのデータは，特許公開公報と特許公報の特許請求の範囲のペアを書き換えの事例として用いることで構築されている。

### 3.2 事前学習済み言語モデルの微調整

本研究では，特許請求の範囲の書き換え生成モデルを事前学習済みの大規模言語モデル [6, 7] を微調整することで実現する。本研究では，特に，日本語 GPT モデルを特許請求の範囲書き換え生成モデルの学習に用いる。より具体的には，言語モデルのパラメタ $\theta$ を 4.1 節の目的関数を最小化するように構築したデータセットを用いて微調整する。本研究が扱う特許請求の範囲の書き換えタスクは，本質的には系列変換タスク（要約，翻訳，質問応答，etc.）に相当する。ここで，特許出願時の特許請求の範囲は生成の際に最初に利用されるコンテキスト情報であり，モデルに対する先行入力（Prompting）として与えて，後続のテキスト（特許公報の特許請求の範囲）を生成するように学習する。

## 4 評価実験設定

本節では，大規模言語モデルに基づいた特許請求の範囲の書き換え生成モデルの学習・評価に用いたデータセット，比較に用いたモデルの概要と学習設定，評価方法について述べる。

### 4.1 学習・検証・評価用データセット

3 節で構築したデータセットから，公開特許公報/特許公報の特許請求の範囲が変化している場合（誤字などの軽微な修正も含む）の全てのペアを抽出し，言語モデルの学習用（1,142,640 件）・検証用（1,000 件）・評価用（1,000 件）データとして用いた。ここで，評価実験に用いる言語モデルが想定する最大トークン長には制約があるため，特許公開公報と特許の特許請求の範囲がそれぞれ 800 トークン長以内に収まるデータのみを用いていた。

### 4.2 比較モデル

本研究では，ヒューリスティックなベースラインモデルと，大規模言語モデルとして日本語 GPT モデルに基づいた書き換え生成モデルの性能を検証した。検証に用いたモデルは以下の通りである。

- **Copy**: 特許請求の範囲を書き換えない場合。
- **RDC** (Random delete of claims): 請求項 1 を除く請求項をランダムに一つ削除する場合。ここで，削除した請求項に従属する請求項も全て除外する。
- **DMMC** (Delete of multi-multi claims): マルチマルチクレームを全て除外する場合。ここで，削除したマルチマルチクレームに従属する請求項も全て除外する。
- **Open-calm-medium**: 日本語 GPT モデル（400M パラメタサイズ）の微調整。
- **Open-calm-large**: 日本語 GPT モデル（830M パラメタサイズ）の微調整。

ここで，RDC は特許の登録査定時の請求項数が出願時と比較して減少する傾向があるという知見に基づいている（3 節）。DMMC は日本国を含む各国特許庁におけるマルチマルチクレームの制限に対する動きを反映したモデルである。open-calm-*はサイバーエージェント社が公開の日本語 GPT モデルである。これらのモデルは，huggingface のレポジトリ[4)]にアップロードされているモデルを用いた。言語モデルのパラメタサイズ $|\theta|$ については，学習コスト，推論速度の観点から扱いやすいパラメタサイズが 10 億（1B）程度までのモデルを用いた。ここで，7B, 13B 相当のモデルであったとしても，元の言語モデルのパラメタ行列の低ランク近似を用いた QLoRA のような手法 [8] を導入することで学習の実施は可能であるが，目標タスクの適応においては元の言語モデルそのものの性能や事前学習に用いられたデータのドメインに非常に強く依存するため，本研究では除外した。

### 4.3 評価指標

特許法第 17 条において定められる特許の補正要件規定を鑑みると，BLEU のような評価指標では，特許請求の範囲を仮に書き換えないでそのままコ

4） http:// https://huggingface.co/

ピーして出力するような保守的なモデルであったとしてもスコアが過大に評価される可能性がある [9]。特許請求の範囲の補正には，請求項の削除や追加，内的付加，外的付加のような補正形態がある。これらの補正形態に基づいた特許請求項の書き換え操作には，基本的には，元の特許請求の範囲の文からのコピー，追加，削除，変更のような異なる種類の書き換え操作が含まれる。こうした，異なる種類の書き換え操作を考慮したモデル出力（書き換え結果）な評価をするためには単純な BLEU スコアの適用は望ましくない。そこで，本研究では 3 節での分析と同様に，異なる編集単位に基づいた編集距離（レーヴェンシュタイン距離）に基づいて参照文と生成文の非類似度（距離）を評価する。用いたスコアは以下のとおりである。

- **line-dist**: 行単位の編集操作を適用した場合の参照文と生成文の編集距離の平均。
- **word-dist**: 単語単位の編集操作を適用した場合の参照文と生成文の編集距離の平均。

ここで，各スコアを計算する際に用いた書き換え生成モデルの出力には，ランダムサンプリング（top_k=50, top_p=0.95）による 50 個の生成候補からスコアが最も最大となる生成文をそれぞれ選択した。

**表 2**　各特許請求の範囲の書き換え生成モデルの評価セットにおける性能

| モデル | $\|\theta\|$ | line-dist | word-dist |
|---|---|---|---|
| Copy | - | 9.98 | 420.7 |
| RDC | - | 10.66 | 565.94 |
| DMMC | - | 10.21 | 473.79 |
| open-calm-medium | 400M | 9.49 | 204.20 |
| open-calm-large | 830M | **8.53** | **170.81** |

## 5　評価実験結果

各特許請求の範囲の書き換え生成モデルの評価セットにおける性能（line-dist，word-dist）を表 2 に示す。結果は，元の特許請求の範囲元からそのまま文をコピーするような保守的なモデルである Copy の書き換え性能が，元の特許請求の範囲から請求項を削除するようなモデルである RDC，DMMC を上回る結果となった。請求項の削除は，特許請求の範囲の補正方法として比較的一般的である一方で，ランダム削除や一括削除による基準では，本来削除するべきでない請求項の情報が欠落した可能性がある。ただし，マルチマルチクレームのみを除去する DMMC は RDC よりも高い性能を示しており，請求項の妥当性を明確な基準を用いて考慮することの重要性を示唆している。大規模言語モデルの微調整に基づいた書き換え生成モデルについについては，パラメタサイズによらずヒューリスティックなベースラインモデルを上回ることが確認できた。特に，単語単位の編集操作に基づいた word-dist で顕著な改善が確認できた。また，言語モデルのパラメタサイズを大きくすることで，特許請求の範囲の書き換え性能が向上することが示された。しかしながら，特許請求の範囲の全文のペアを生成するように学習を行う現状のモデルは，モデルのパラメタサイズの増大によりモデルの学習に必要となる計算リソースも飛躍的に増大するという制約がある。

## 6　まとめと今後の課題

本研究では，特許請求の範囲の書き換えを自動生成することを目的とした書き換え生成モデルの基礎的検討を実施した。具体的には，特許請求の範囲の書き換えタスクの定義，書き換え生成モデルの学習・評価のためのデータセット構築，大規模言語モデルに基づいた提案モデルの性能とその限界を調査した。評価実験では，大規模言語モデルに基づいた提案モデルがヒューリスティックなベースラインモデルを凌駕する性能を示した。また，事前学習モデルのパラメタサイズがモデルの書き換え性能に影響を与えることが示唆された。今後の課題としては，提案モデルの性能をさらに向上するために，より大規模な言語モデルを本研究タスクで微調整や，書き換え対象の特許請求の範囲だけではなく，先行特許の情報や予想される特許の拒絶理由等を書き換えモデルに対する追加の入力情報として用いることを検討する。また，自動評価だけではなく人間の専門家による人手評価を実施する。

## 謝辞

本研究は，JST 戦略的創造研究推進事業 ACT-X（JPMJAX22A4），ムーンショット型開発支援事業（JPMJMS2236），JSPP 科研費基盤 C（19K12116）の支援により実施された。

## 参考文献

[1] Mihai Lupu, Katja Mayer, Noriko Kando, and Anthony J Trippe. **Current challenges in patent information retrieval**, Vol. 37. Springer, 2017.

[2] Atsushi Fujii, Makoto Iwayama, and Noriko Kando. Overview of the patent retrieval task at the ntcir-6 workshop. In **NTCIR**, 2007.

[3] Akihiro Shinmori, Manabu Okumura, Yuzo Marukawa, and Makoto Iwayama. Patent claim processing for readability-structure analysis and term explanation. In **Proc. of the ACL-2003 workshop on Patent corpus processing**, pp. 56–65, 2003.

[4] 新森昭宏, 大屋由香里, 谷川英和ほか. 特許請求項における多重多数項引用の検出と書き換え. 情報処理学会論文誌, Vol. 49, No. 7, pp. 2692–2702, 2008.

[5] Akira Goto and Kazuyuki Motohashi. Construction of a japanese patent database and a first look at japanese patenting activities. **Research Policy**, Vol. 36, No. 9, pp. 1431–1442, 2007.

[6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. **Advances in neural information processing systems**, Vol. 30, , 2017.

[7] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. **OpenAI blog**, Vol. 1, No. 8, p. 9, 2019.

[8] Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. Qlora: Efficient finetuning of quantized llms. **arXiv preprint arXiv:2305.14314**, 2023.

[9] 河野誠也, 野中尋史, 吉野幸一郎. 大規模言語モデルに基づいた特許請求の範囲の自動書き換え生成モデル. 研究報告自然言語処理 (NL), Vol. 2023-NL-258, No. 24, pp. 1–6, 2023.

[10] Jieh-Sheng Lee and Jieh Hsiang. Patent claim generation by fine-tuning openai gpt-2. **World Patent Information**, Vol. 62, p. 101983, 2020.

[11] Jieh-Sheng Lee. Evaluating generative patent language models. **World Patent Information**, Vol. 72, p. 102173, 2023.

# A　付録

## A.1　関連研究

Transformer 言語モデルに基づいた大規模言語モデルの登場により，文章要約，質問応答，機械翻訳，文法誤り訂正といった，様々な自然言語処理のタスクで，大規模言語モデルを活用した手法が最先端の性能を示している [6]。大規模言語モデルの性能向上に伴いこうした応用タスクの性能も進化を続けている一方で，一般のコーパスとは語彙が異なる特許ドメイン（特に米国特許）に特化した特許言語モデルの開発と特許ドメインにおける様々なタスクへの応用が進んでいる [10, 11]。こうした特許言語モデルの性能評価では，しばしば特許請求項の自動生成タスクに焦点があてられるが，与えられた請求項に対する実用的な書き換えを生成するようなタスクも存在する [4]。請求項の書き換えタスクでは，単純な生成タスクとは異なり，元々の請求項の意味内容を概ね保持しつつ，その品質を向上させた書き換えを生成する必要がある。ここで，書き換えの品質は，単に言語モデル的に自然であるということではなく，特許が拒絶されない，より強い権利範囲を持つ，といった特許制度上の観点から決定されることに注意されたい。請求項の書き換えタスクに関しては，ルールベース手法に基づいた多重多数項引用の請求項（マルチマルチクレーム）を複数の請求項に分解して書き換えるような研究が存在するが，非常に限定された目的のみでの書き換え支援にとどまっている [4]。また，単一の請求項だけでなく特許請求の範囲全体の構造を考慮した一貫的な書き換えを自動生成するような研究は存在しない。また，そのような書き換え生成モデルを学習・評価するためのデータセットも存在しない。そこで，本研究では多目的に対応した特許請求の範囲の書き換えを自動生成するモデルを実現するためのデータセットの構築と大規模言語モデルの応用可能性を検討する。

## A.2　特許出願における公報発行プロセス

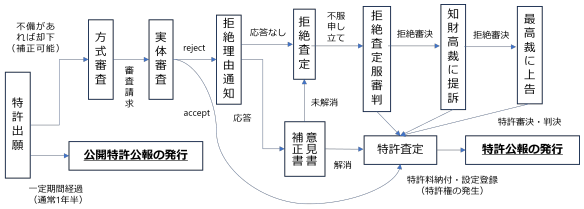


**図 1**　特許出願における特許公開公報と特許公報の発行までのプロセス

## A.3　実際の書き換え事例

特許請求の範囲の書き換え事例[5)]を以下に示す。この書き換えの例では，請求項 1 の記載を減縮し，請求項 3 の記載においては，訂正前の請求項 1 の記載を組み込んで，請求項 1 の記載を引用しないものとし，独立形式請求項として改めている。また，請求項 4 の記載においては，請求項 2 の記載と同様に独立形式請求項へ改めるとともに，請求項 4 の記載を減縮した事例である。

**【変更前】**

1. 【請求項 1 】
2. 回転支持部により回動自在である一対の鋏片のうち、
3. 第一の鋏片の刃縁にＡ形状を有し、
4. 第二の鋏片の刃縁にＢ形状を有することを特徴とする、鋏。
5. 【請求項 2 】
6. 前記第二の鋏片のＢ形状の厚みがＤの範囲であることを特徴とする、請求項 1 に記載の鋏。
7. 【請求項 3 】
8. 前記第一及び第二の鋏片の刃先部に更にＣ形状を有することを特徴とする、請求項 1 に記載の鋏。
9. 【請求項 4 】
10. 前記第一及び第二の鋏片の刃部の素材がＥであることを特徴とする、請求項 1 に記載の鋏。

**【変更後】**

1. 【請求項 1 】
2. 回転支持部により回動自在である一対の鋏片のうち、
3. 第一の鋏片の刃縁にＡ形状を有し、
4. 第二の鋏片の刃縁に凸略円弧と凹略円弧が交互に連続したＢ形状を有することを特徴とする、鋏。
5. 【請求項 2 】
6. 前記第二の鋏片のＢ形状の厚みがＤの範囲であることを特徴とする、請求項 1 に記載の鋏。
7. 【請求項 3 】
8. 回転支持部により回動自在である一対の鋏片のうち、
9. 第一の鋏片の刃縁にＡ形状を有し、
10. 第二の鋏片の刃縁にＢ形状を有し、
11. 前記第一及び第二の鋏片の刃先部に更にＣ形状を有することを特徴とする、鋏。
12. 【請求項 4 】
13. 回転支持部により回動自在である一対の鋏片のうち、
14. 第一の鋏片の刃縁にＡ形状を有し、
15. 第二の鋏片の刃縁にＢ形状を有し、
16. 前記第一及び第二の鋏片の刃部の素材がＦ成分を含むＥであることを特徴とする、鋏。

5)　特許庁「訂正審判請求書及び訂正請求書の記載例」より引用

## A.4　拒絶理由毎の平均編集距離

表 3 は，特許出願中に発送された拒絶理由毎に書き換えペア間の平均編集距離を計算した結果である。ここで，line-dist，word-dist はそれぞれ編集操作に行単位と単語単位を用いる場合に相当する。また，表 3 に用いた書き換えペアは，構築したデータセットのサブセットであり，2021 年に出願され，2022 年までに特許査定を受けたもののうち拒絶理由（2021 年に改訂された拒絶理由条文コードに基づく）の発送履歴があるものに限定していることに注意されたい。

**表 3**　拒絶理由毎の書き換えペアの平均編集距離

| 拒絶理由 | line-dist | word-dist |
|---|---|---|
| 第 29 条第 1 項第 3 号—頒布された刊行物に記載された発明 | 19.87 | 319.55 |
| 第 29 条第 2 項—進歩性 | 19.26 | 316.97 |
| 第 36 条第 6 項第 2 号—明確性要件 | 18.47 | 283.58 |
| 第 36 条第 6 項第 1 号—サポート要件 | 20.48 | 269.33 |
| 第 36 条第 4 項第 1 号 (実施可能要件)—実施可能要件 | 16.20 | 265.95 |
| 第 36 条第 4 項第 1 号 (委任省令要件)—委任省令要件 | 22.64 | 468.16 |
| 第 29 条の 2—拡大先願 | 19.49 | 322.16 |
| 第 37 条—発明の単一性 | 27.84 | 519.34 |
| 第 29 条第 1 項柱書 (非発明)—発明該当性 | 17.24 | 279.12 |
| 第 29 条第 1 項第 2 号—公然実施をされた発明 | 20.43 | 459.50 |
| 第 39 条第 2 項—協議が成立せず、又は協議をすることができないとき | 18.68 | 438.65 |
| 第 17 条の 2 第 3 項—新規事項を追加する補正 | 26.59 | 489.37 |
| 第 17 条の 2 第 4 項—発明の特別な技術的特徴を変更する補正 | 18.50 | 257.50 |
| 第 29 条第 1 項柱書 (非利用)—産業上の利用可能性 | 23.90 | 290.50 |
| 第 39 条第 1 項—異なった日の特許出願、最先の特許出願 | 14.96 | 421.60 |
| 第 36 条第 6 項第 3 号—簡潔性要件 | 8.00 | 301.00 |
| 第 36 条第 6 項第 4 号—特許請求の範囲の記載に関する委任省令要件 | 16.45 | 229.73 |
| 第 36 条第 4 項第 2 号—先行技術文献情報開示要件 | 18.00 | 197.00 |

## A.5　学習設定

言語モデルの学習には RTX A6000 を 4 枚用いた。さらに，学習に用いる言語モデルのサイズに応じて，各 GPU デバイスにおけるバッチサイズ，gradient accumulation steps を実質的な学習バッチサイズが 512 以上となるように設定した。モデルの Optimizer には，学習率 5e-5 の AdamW を設定した。モデルの学習エポック数は最大 1 として，0.1 エポック毎にチェックポイントを設定した。また，学習・推論時においては，浮動小数点スケール（bfloat16）に基づいた量子化を適用した。モデルの評価には検証セットにおいて最も低い Perplexity を示したチェックポイントを用いた。