

法令文解析に適した事前学習モデルの構築

渋谷太朗¹ 中村誠¹¹新潟工科大学 工学部202111079@cc.niit.ac.jp mnakamur@niit.ac.jp

概要

比較法学研究を推進するために日本法と外国法の類似条項を対応付ける研究が行われている。その手法の一つに汎用言語モデルである BERT が使われている [1]。これにより、事前学習と fine-tuning の組み合わせで End-to-end 学習を可能とする。法令分野の事前学習モデルは既にいくつかある [2, 7] が、判例など法令以外の文書もコーパスに含まれており、法令のみをコーパスとした BERT モデルは存在せず、高精度な法令解析を行うには、法令のみをコーパスとした BERT モデルが必要となる。したがって、本研究の目的は、日本の法令文解析に適した事前学習モデルを構築することである。法令のみをコーパスとした事前学習モデルを二通り作成し、各モデルの言語理解能力、外国法対応付け能力をテストした。結果として、F1 値で東北大学が発表した、日本語汎用事前学習モデル (以降、東北大 BERTⁱ) を上回る性能を示した。

1 はじめに

比較法学とは、種々の法体系における制度又は法の機能を比較することを目的とする学問である。比較法学では原則として国家の法体系を対象として比較を行う。このとき、法を原文のまま理解し、相違点と共通点の発見を経て国家間の法制度や法文化を比較する [3]。比較法学の実務的な効用として、自国法の立法整備、解釈・適用の改善が挙げられる。法制審議会では、諸外国の法制度についてまとめられた資料や制度の有無を比較した表が度々参照されている。また、グローバル化の進む現在、比較法は法学において重要な役割を担っており、海外の文化や考え方を理解するための材料にもなり得る [3]。

BERT (Bidirectional Encoder Representations from Transformers) とは、Google AI が 2018 年に発表し

ⁱ<https://github.com/cl-tohoku/bert-japanese>

ⁱⁱ<https://www.e-gov.go.jp/>

た深層学習モデルである。BERT は事前学習と fine-tuning の組み合わせでモデルの構築を行う [1]。本研究では、事前学習のみに焦点を当てている。

計算機による類似条文の対応付けの自動化は、条文を比較することにより可能となる。先行研究 [4] で BERT が採用されたが、事前学習済みモデルを使用していたため、日本法同士の対応付けでは Jaccard 係数や BM25 といった単語ベースの手法に比べて劣る性能となった。一般的な BERT モデルは Wikipedia などのテキストで訓練されており、医療等の特定ドメインに対しては性能が低いことが報告されている [5]。法令は一般文書より硬い表現で書かれるため、法令文を扱うタスクを行う場合、BERT モデルの法令ドメイン適応が必要となる [6]。法令ドメインに適応した BERT モデルは存在する [2, 7] が、コーパスに判例文などの法令以外の文書も含まれており、高精度な解析を行うには、法令のみをコーパスとした BERT モデルが必要となる。

したがって、本研究の目的は、日本の法令文解析に適した事前学習モデルを構築することである。モデル構築後、言語理解能力と外国法対応付け能力をテストし、法令分野において汎用言語モデルを上回る性能を示すと共に、法令タスクについてもこれを上回る性能であることを示す。

本稿は以下のような構成になっている。第 2 節でモデルの構築を述べる。第 3 節で実験の目的と設定、評価タスクを述べる。第 4 節で各実験結果を示し、考察を行う。第 5 節で本稿のまとめと今後の課題について述べる。

2 モデルの構築

2.1 法令データの収集

本稿において、法令とは、憲法、法律、政令、府令、省令、勅令の総称である。日本法令は e-Govⁱⁱにて公開されており、XML, HTML, RTF, PDF の形式でダウンロードする事が可能である。本研究では、コーパスを作成する際に、XML 形式で法令をダウンロ

ードし、解析を行った。

日本の法令 XML の構造は、法令標準 XML スキーマファイルⁱⁱに沿って構築されている。法律名は<LawTitle>、法律番号は<LawNum>、各文は<Sentence>など、各要素で固有のタグが指定されている。図 1 に例を示す。法令 XML ファイルを解析する事で、膨大な法律文書から、任意の指定した情報のみを抽出する事が可能となる。

```
<LawNum>昭和二十四年法律第二百六十八号</LawNum>
<LawBody>
  <LawTitle Kana="ぎょうしこうほう"Abbrev=""AbbreKana="">漁業法</LawTitle>
  <MainProvision Extract="true">
    <Article Num="1">
      <ArticleCaption>(現行漁業法の存続)</ArticleCaption>
      <Article Title>第一条</Article Title>
      <Paragraph Num="1">
        <Paragraph Num/>
        <Paragraph Sentence>
          <Sentence Function="main" Num="1">漁業法(昭和二十四年法律第二百六十八号。
          <Sentence Function="main" Num="2">ただし、新法大六十七条の規定及び同条に係
```

図 1 法令 XML の例

2.2 コーパス

一般に、ドメイン適応するモデルは、コーパスにドメイン以外の方法を含めるとノイズとなり、学習効率の悪化だけでなく、学習時間の増加にもつながる。また、事前学習時点でクリーンなコーパスを用いないと、fine-tuning での学習効果が薄く、適切に学習がされないことが報告されている [8]。コーパスに含めるべき情報を選定するため、先行研究 [9] に従い、<LawTitle>、<LawNum>、<Sentence>の三種のタグを対象として、コーパスを作成した。作成したコーパスの詳細を以下に示す。

法令コーパス 日本の法令 8,466 本(令和 6 年 12 月 5 日時点)を対象。文書数は 2,507,874、一文当たりの平均文字数は約 54.9 字、コーパスサイズは約 417[MB]である。

2.3 用いたパラメータ

各モデルで用いたパラメータは共通値である。パラメータを表 1 に示す。これらのパラメータは先行研究 [1] と NDIVIA 公式の GitHubⁱⁱⁱに公開されている BERT の事前学習や fine-tuning の実装例を参考に決定した。また、モデルの構造は BERT_{base} [1] である。

表 1 各モデルのパラメータ

(1)エポック数	10	(4)学習率	5e-5
(2)バッチサイズ	16	(5)MLM 確率	0.15
(3)FP16	True	(6)NSP 確率	0.50

ⁱⁱⁱ<https://laws.e-gov.go.jp/docs/law-data-basic/419a603-xml-schema-for-japanese-law/>

2.4 BERT の事前学習タスク

一般に BERT の事前学習タスクは Masked Language Model (MLM) と Next Sentence Prediction (NSP) の二種類が用いられている。BERT は事前学習時に MLM と NSP を同時に実行する [1]。以下に各タスクの特徴を示す。

MLM タスク 文中のトークンを一定確率でマスクトークンに置き換え、マスクされたトークンをモデルが予測するタスクである。図 2 に例を示す。

MLM 確率でマスクするトークンの割合を決定し、一般的に 0.15 (全体の 15%) と設定される。マスクトークンは、80%でマスクトークン、10%で他のランダムなトークン、10%で元単語に置き換わる。MLM タスクを通じて、モデルは文脈内での個々の単語の意味を学習する [1]。

(1) "スキー場"に"[MASK]"が"積"もる。" 回答:雪

(2) "ChatGPT"は"生成AI"[MASK]"一つ"だ。" 回答:の

図 2 MLM タスクの例

NSP タスク 文 A と文 B があり、各文が連続している文か、連続していない文かをモデルが予測するタスクである。図 3 に例を示す。NSP 確率で連続している文 (Positive Example)、連続していない文 (Negative Example) の割合を決定し、一般に 0.5 と設定される。NSP タスクを通じて、モデルは文間の関係性を理解する [1]。

(Positive Example) 文A:スキー場に雪が積もった。 正解:○
文B:明日はスキーに行こう。

(Negative Example) 文A:スキー場に雪が積もった。 正解:×
文B:明日の朝食はパンだ。

図 3 NSP タスクの例

2.4 作成した事前学習モデル

本実験で作成した事前学習モデルの詳細を以下に示す。また、実行環境は、NDIVIA Geforce4090、メインメモリ 64[GB]を用いた GPGPU 環境である。

法令 BERT_MLM 法令コーパス，表 1 のパラメータ，事前学習タスク MLM でモデルを作成した．モデルの作成時間は約 23 時間である．

法令 BERT_MLM+NSP 法令コーパス，表 1 のパラメータ，事前学習タスク MLM+NSP でモデルを作成した．モデルの作成時間は約 288 時間である．

3 実験

3.1 実験方法

本稿では以下の二つの実験を行った．

実験 1: 事前学習タスクの選別実験

実験 2: 日本法と外国法の対応付け実験

実験 1 では，最適な事前学習タスクを決定する実験を行った．コーパス，パラメータを固定し，異なる事前学習タスクを二種類用いて，モデル作成を行った．モデル作成後，3.2 節に示す MLM タスクで各モデルの性能をテストした．

実験 2 では，各モデルで日本法と外国法を対応付ける実験を行った．各モデルで，3.3 節に示す日韓法対応付けタスクで各モデルの性能をテストした．

3.2 MLM タスク

問題数を 100 問とし，モデルの確信度一位の単語を正解と比較する，Accuracy で評価を行った．また，前処理として「、」や「。」などの特殊記号が正解にならないようにし，McCab で形態要素解析をした後，問題文のシーケンス長を 10~512 に制限した．オープンテストとクローズドテストを作成するため，e-Govⁱ よりダウンロードした法令文を問題文として一つ，Wikipedia よりダウンロードしたテキスト文を問題文として一つの合計二つのタスクを作成した．各タスクの特徴を下に示す．

実法令 e-Govⁱ より，電気事業法（昭和 39 年第 170 号），ガス事業法（昭和 29 年第 51 号），母体保護法（昭和 23 年第 156 号）をダウンロードし，問題文を作成した．モデルは日本の現行法令をコーパスとしているため，本タスクはクローズドテストとなる．

Wikipedia_法律 Wikipedia より法律カテゴリに属するページを 100 ページダウンロードした．モデルのコーパスに Wikipedia の文書は含まれていないので，本タスクは オープンテストとなる．

3.3 日韓法対応付けタスク

日本法と外国法の対応付けタスクとは，任意の日本法と，それに類似する外国法を比べ，類似条文をモデルに予測させるタスクである．このタスクにより，モデルの外国法対応付け能力をテストすることができる．本タスクの対応付ける法律は，日本の船員法（昭和 22 年法律 100 号）と日本語訳された韓国の船員法である．対応付けの例を図 4 に示す．韓国語は，日本語と構造が近く，機械翻訳がしやすい．また，日本法と韓国法は法域が近く，人手による正解データの作成も容易である [11]．問題数を 100 問とし，TP，TN，FP，FN から Accuracy，Precision，Recall を算出し，F1 値を求めた．

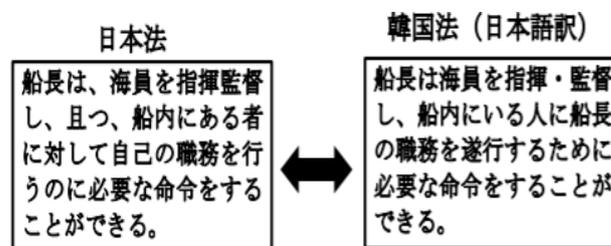


図 4 対応付けされた類似条文（船員法） [11]

3.4 実験 1: 事前学習タスク選別実験

本実験では，モデルの言語理解能力を測定する事を目的とする．2.4 節に示す二つのモデルに，3.2 節に示す MLM タスクを実行し，モデルの言語理解能力をテストした．

3.5 実験 2: 日韓法対応付けタスク

本実験では，モデルの外国法対応付け能力を測定する事を目的とする．2.4 節に示す二つのモデルに，3.3 節に示す日韓法対応付けタスクを実行し，モデルの外国法対応付け能力をテストした．

4 実験結果と考察

4.1 実験 1: 事前学習タスク選別実験

実験結果を表 2 に示す．クローズドテストである実法令では，(2) が最高値を示しており，オープンテストでは，(3) が最高値を示す結果となった．(2) は，タスクが法令文書の場合，高い言語理解能力を発揮するが，法令分野であっても，Wikipedia などの法令文書以外がタスクの場合，言語理解能力は (3) モデルの方が高いことを示した．

(1) と (2) では、オープンテスト、クローズドテスト共に、(2) が上回った。BERT のその後の研究で、適切に MLM タスクを実行できていれば、NSP タスクによる追加の学習効果は薄いことが示されており [10]、逆説的に、(1) は、適切な MLM タスクを実行できていない可能性がある。

4.2 実験 2: 日韓法対応付けタスク

実験結果を表 3 に示す。Precision 以外の値において、(1) と (2) が同値で最高値を示す結果となった。Accuracy と Recall が高いという結果より、TP と TN が高いと予測できる。また、Precision が低いことから、FP が多いと予測できる。したがって、(1) と (2) は、(3) に比べ、正答率は 4% 高いが、誤回答率も 4% 高いと読み取れる。

表 2 の MLM タスクでは、(1) と比べ、(2) が大差で高性能を示していた。MLM タスクは、モデルの言語理解能力をテストするタスクであり、NSP タスクによる事前学習の性能向上が確認できた。しかし、表 3 の日韓法対応付けタスクでは、NSP タスクによる事前学習の性能向上が確認できなかった。この結果より、本タスクにおいて、モデルの言語理解能力はあまり重要視されないと考察できる。

一般に、深層学習モデルは、メモリ効率、計算コスト、保存領域などのハードウェア観点より、軽量モデルが望まれる。F1 値を見ると、(1) と (2) が、同値であり、(1) と、(2) は、モデル作成時間が約 12.5 倍であった。以上の結果を総合的に判断すると、性能値が同値であり、モデルサイズが軽量である (1) が本タスクにおいて、総合的に優れていると言える。

表 2 各モデルの MLM タスク評価値

	実法令	Wikipedia_法律
(1)法令BERT_MLM	29	9
(2)法令BERT_MLM+NSP	80	36
(3)東北大BERT	66	58

表 3 各モデルの日韓法対応付けタスク評価値

	Accuracy	Precision	Recall	F1値
(1)法令BERT_MLM	0.45	0.78	0.42	0.54
(2)法令BERT_MLM+NSP	0.45	0.78	0.42	0.54
(3)東北大BERT	0.41	0.82	0.34	0.45

5 おわりに

本研究では、日本の法令文解析に適した事前学習モデルを構築し、MLM タスクと、日韓法対応付けタスクを行い、モデルの言語理解能力と外国法対応付け能力を調査する事を目的とした。結果として、両タスクにおいて、一部東北大 BERT を上回る性能を示した。また、MLM タスクにおける、NSP タスクによる事前学習の重要性と日韓法対応付けタスクにおける、NSP タスクによる事前学習の不要性を示した。今後は、韓国以外で日本法との対応付けタスクを作成し、韓国法以外でのモデルの性能検証、事前学習タスクである MLM タスクの設定を見直し、モデルの性能向上が達成されるかの検証、モデルの fine-tuning を行い、更なる性能向上を目指せるかの検証を行って行きたい。

謝辞

本研究は JSPS 科研費 JP24K03230 の助成を受けたものです。

参考文献

- [1] J. Devlin, M.W. Chan, K. Lee, K. Toutanova. “BERT:Pre-training of deep bidirectional transformers for language understanding.” In *NAACL-HLT*. pp.4171-4186 (2018)
- [2] 宮崎, 菅原, 山田, 徳永. 日本語法律分野文書に特化した BERT の構築. *NLP2022*, pp1546-1551 (2022)
- [3] 滝沢. 比較法. 三省堂 (2020)
- [4] H. Cho, R. Koseki, A. Shima, M. Nakamura. “Mapping Similar Provisions Between Japanese and Foreign Laws” In *New Frontiers in Artificial Intelligence*, pp.36-50 (2023)
- [5] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C.H. So, J. Kang. “BioBERT: a pre-trained biomedical language representation model for biomedical text mining.” *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240 (2020)
- [6] D. Yamada, M. Nakamura. “Development of a Multilingual Model for Mapping Japanese and Foreign Laws.” In *JURISIN2024*, pp.128-141 (2024)
- [7] 星野, 狩野. 司法試験自動回答を題材にした BERT による法律分野の含意関係認識. *NLP2020*, pp.577-580 (2020)

- [8] V. Ramanujan, T. Nguyen, S. Oh, L. Schmidt, A. Farhadi. “On the Connection between Pre-training Data Diversity and Fine-tuning Robustness.” In *NeurIPS2023* (2023)
- [9] 渋谷, 中村. 異なるコーパスを用いた事前学習モデルによる穴埋め問題の比較.IEICE2024, p.50 (2024)
- [10] 黒橋, 自然言語処理〔三訂版〕, 放送大学教育振興会 (2023)
- [11] 高橋, 中村. 異なる法域と法律による類似条項の自動対応付け比較. IEICE2024. p.60 (2024)