

法律間の類似条文の対応付けにおける BERT の法令ドメイン適応

山田大地¹ 中村誠¹¹新潟工科大学 工学部

202011233@cc.niit.ac.jp mnakamur@niit.ac.jp

概要

比較法研究では日本法と外国法の類似する条文を対応付けることがある。既に BERT [1] を用いて条文を対応付ける研究がなされているが、モデルは事前学習のみで使われていた。本研究では、BERT の fine-tuning を行い、条文を対応付けてその有効性を検証することを目的とする。fine-tuning には教師なし学習と教師有り学習を用いて性能の比較を行った。また、試験的に多言語モデルを用いた対応付け結果を示す。実験より、法令テキストを用いて fine-tuning を行ったモデルは従来の手法を上回り、特に教師あり学習が類似条文の対応付けにおいて最も高い性能を発揮した。

1 はじめに

比較法とは種々の法体系における法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。比較法では原則として国家の法体系を対象として比較を行う。通常、比較法では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 [2]。比較法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。

比較法は日本法と外国法の類似点を足掛かりとして行う場合がある。正確な対応付けには法的な知識や使用言語への理解が求められ、そのような人材は一部の専門家に限られる。このとき、日本法と外国法の類似部分を自動で対応付けることが出来れば、比較法研究に寄与するとともに、一般にも海外ビジネスを始める際に有用である。

計算機による類似部分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り、既に長、小関らがやっている [3,4,5]。提案手法の一つに 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 [6]。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。

そこで、本研究では BERT モデルの fine-tuning を Transformer-based Sequential Denoising Auto-Encoder (TSDAE) を用いた教師なし学習と、ラベル付きデータを用いた教師あり学習の 2 種類で行い、法律間の類似条文の対応付けにおける fine-tuning の有効性を検証することを目的とする。また、日本法と外国法の対応付けについて、多言語モデルが有効かどうか調査する。

本稿は以下のような構成となっている。第 2 章で関連研究のついて述べる。第 3 章で日本法同士の対応付け、日本法と外国法との対応付けの 2 種類の実験を説明する。第 4 章で実験結果を示し、考察を行う。第 5 章で本稿のまとめと言語の課題について述べる。

2 関連研究

2.1 類似条文の対応付け

長、小関ら [3,4,5] は、類似した条文の対応付けについて、条文を一つの文書とした類似文書検索と捉え、図 1 のように行った。初めに、異なる法律の

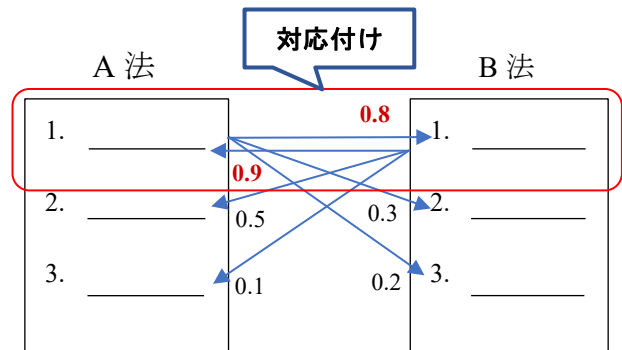


図 1 条文の対応付け

条文の組み合わせ全てに対し、類似度を計算する。次に、ある条文に対して最も類似度が高い条文を参照し、その条文から見て最も類似度が高かった条文が元の条文と一致した場合に2つの条文を対応付けている。

対応付けに用いる文書間の類似度に関しても様々なものが試されている。まず、文書を単語の集合として扱い、類似度算出には Jaccard 係数と Dice 係数が用いられた [3]。しかし、これらの手法は単語の重要度を考慮せず、類義語であっても全く別の単語とする等の欠点があった。そこで、BERT による埋め込み表現の cosine 類似度を用いて対応付けが行われた [4]。その他、単語の重要度をもとに文書の順位付けを行う BM25、入力長制限の無い Doc2Vec についても試されている [5]。

2.2 Sentence-BERT

Sentence-BERT (SBERT) とは、事前学習済みの BERT モデルを Siamese Network を用いて fine-tuning を行ったモデルであり、従来の BERT に比べて文埋め込み表現が改善されるほか、複数の文書を扱う際に効果を発揮する [7]。そこで用いられた目的関数のうち、Regression Objective Function を取り上げる。初めに、モデルは二つの文を受け取り、文埋め込みの cosine 類似度を計算する。次に正解ラベル（類似度）との平均二乗誤差を計算し、似ている文書に対しては類似度が大きくなるよう埋め込みを調整する。

2.3 TSDAE

TSDAE とは教師なし文埋め込みを学習するための Transformer モデルである [8]。図 2 に TSDAE の構造を示す。TSDAE は、入力文にノイズ（トークンの削除）を加えて破損させた文を固定長のベクトルにし、そのベクトルを元の入力文に再構成することで文埋め込みを学習する。Wang らは、入力トークンの削除率が 0.6 で最適であると示した [8]。

既存の Transformer モデル [9] とは、Cross-Attention のキーとバリューが文埋め込みに限定される点で異なる。Cross-Attention は以下の式で計算される。

$$H^{(k)} = \text{Attention}(H^{(k-1)}, [s^T], [s^T]) \quad (1)$$

ⁱ いずれも令和 3 年 2 月時点。

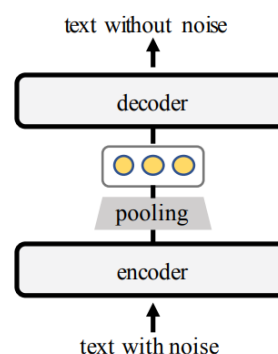


図 2 TSDAE の構造 [8]より引用

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (2)$$

ここで、 $H^{(k)} \in \mathbb{R}^{t \times d}$ は k 番目の層における t 回目の decoding ステップのデコーダの隠れ状態、d は文埋め込みのサイズ、 $[s^T] \in \mathbb{R}^{1 \times d}$ は文埋め込みを含む 1 行の行列、Q, K, V はそれぞれクエリ、キー、バリューを表す。

3 実験

3.1 実験方法

本稿では以下のような 2 つの実験を行った。

実験 1：日本法同士の対応付け

実験 2：日本法と外国法の対応付け

実験 1 では法令文書を扱うタスクに対する fine-tuning の効果を、教師なし学習・教師あり学習それぞれについて検証する。先行研究 [5] と比較するため、日本法同士である電気事業法（昭和 39 年法律第 170 号）全 315 条とガス事業法（昭和 29 年法律第 51 号）全 207 条の対応付けを行うⁱ。どちらもインフラに関する法律であるため、構成や条文の内容が類似しているといった特徴がある。図 3 にその例を示す。それぞれの法律を条ごとに分割し、一方の法律の条文と類似した他方の条文とを対応付ける。

電気事業法	ガス事業法
第56条 経済産業大臣は、一般用 電気工作物が経済産業省 令で定める技術 基準に 適合していないと認める とき...	第161条 経済産業大臣は、消費機 器が第百五十九条第二項 令で定める の経済産業省令で定める 技術上の基準に適合して いないと認めるとき...

図 3 類似条文の例

実験2では日本民法（明治29年法律第89号）の家族法（第725条～第881条）とドイツ民法の対応付けを行う。先行研究 [5]ではテキストを英語に統一し、単言語のBERTによる対応付けを行っていた。本研究では、多言語モデルを用いてそれぞれの法律を原文のまま対応付け、先行研究 [5]の結果と比較する。

全実験を通してSBERTのpooling処理にはMEANを採用し、各トークンの平均を文埋め込みとした。また、512トークンを超えた箇所は削除した。BERTの他に先行研究 [5]で採用されていたJaccard係数とBM25についても実験を行い、これらを含め性能を比較する。条文の前処理については、McCabを用いて形態素ごとに分割し、辞書にはNeologdを使用した。

3.2 評価方法

実験1の評価には先行研究 [5]と同じ正解データを用いる。この正解データは、法律の専門家による人手で作成されたものである。対応付け結果を正解データと比較してAccuracy, Precision, Recall, F1値を算出した。

実験2では、実験1のような1対1の対応付けだと殆ど対応が取れないことが分かった。そこで、日本法から見たドイツ法の類似度上位5箇条を対応していると見做し、これを正解データと比較する。結果は以下の式で評価される。

$$\text{正答率} = \frac{\text{正解した日本法の条数}}{\text{ドイツ法と対応する日本法の条数}} \quad (3)$$

なお、正解データは新注釈民法 [10]と新版注釈民法 [11,12]から作成されており、ドイツ法と対応する日本法の条数は106条あった。

3.3 実験1:日本法同士の対応付け

3.3.1 教師なし学習

学習データ 教師無し学習では法令テキストのみを用いる。e-Govⁱⁱ法令検索に掲載されている「工業」カテゴリに属する全485法令をxml形式でダウンロードし、項単位でテキストを抽出した（全42,147項）。

ⁱⁱ <https://elaws.e-gov.go.jp/>

ⁱⁱⁱ [cl-tohoku/bert-base-japanese-whole-word-masking](https://github.com/cl-tohoku/bert-base-japanese-whole-word-masking)

実験設定 学習に用いるベースモデルは東北大学より公開の事前学習済みモデルⁱⁱⁱを利用した。バッチサイズは8、学習率は3e-5、エポック数1で学習を行った。

3.3.2 教師あり学習

学習データ 教師あり学習を行うにはSTS benchmark [13]のようなラベル付きのデータセットが必要である。しかし、法令文に適応したデータセットは作成されていない。そこで、本研究では仮の法令STSデータセットを構築してこれを学習に用い、5分割交差検証によりモデルの性能を評価する。本実験で使用する正解データを基に、「電気事業法-ガス事業法」となる条文の組み合わせ全65,205ペアに対し、類似度を表した-1~1のラベルを付与する。条文ペアのうち実際に対応関係があるペアに対しては1とする。対応関係がないペアに対しては、3.3.1節で作成したモデルを用いて計算した類似度を0.95倍にて低減させた数値をラベリングする。

実験設定 学習に用いるベースモデルは3.3.1節と同様のモデルを利用した。バッチサイズは8、学習率は2e-5、エポック数3で学習を行った。5分割検証の際、学習データと検証データの分割は8:2で行い、ランダムにサンプリングしている。最終的に学習データは52,614件、検証データは13,041件となった。

3.4 実験2:日本法と外国法の対応付け

使用データ 対応付けに用いる法律は、それぞれの政府で提供されているデータをxml形式で入手した。

実験設定 現状、正解データが民法の家族法部分でしか作成されていないため、家族法に絞り実験を行う。今回の実験では、多言語モデルで先行研究と同様の結果が得られるかを試験的に確認するため、Hugging Face^{iv}で公開されている以下のモデルをfine-tuningせず使用する。

- sentence-transformers/LaBSE
- sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2
- sentence-transformers/paraphrase-xlm-r-multilingual-v1

^{iv} <https://huggingface.co/models>

これらのモデルで対応付けた結果を先行研究 [5]でも用いられた単言語の BERT を含め比較する。また、モデルは bert-base-uncased を使用した。

4 実験結果と考察

表 1 に各手法の評価結果を示す。事前学習済み BERT を BERT-pretrained, 教師なしで学習したモデルを SBERT-TSDAE, 教師ありで学習したモデルを SBERT-STS とする。SBERT-TSDAE の評価値は 50 回の実験と性能評価を繰り返して平均をとった数値である。また、SBERT-STS は 5 分割交差検証を行った際の評価である。

4.1 日本法同士の対応付け

4.1.1 教師なし学習

初めに従来の 3 つの手法 Jaccard, BM25, BERT-pretrained について言及すると、日本法同士の対応付けでは Jaccard 係数が最も優れていることが分かる。これは、条文が対応関係にある場合に、共通する単語が多数存在することに起因する。ここで BERT モデルに着目すると、法令コーパスを用いて fine-tuning を行った SBERT-TSDAE は F1 値を約 5.2pt 向上させ、BERT-pretrained を有意に上回った ($p < 0.05$)。特に、Recall が大きく改善されており、類似した条文をより多く対応付けることに成功した。

4.2 教師あり学習

SBERT-STS は、全ての手法を上回り、SBERT-TSDAE に比べて F1 値を更に約 2.6pt 向上させた。SBERT-STS と同様 Recall の伸びが見られたが、今回は Precision も伸びており、教師あり学習の方が教師なし学習と比べ優れている結果となった。

今回使用したデータセットは電気事業法とガス事業法のペアのみで作成しているため、データセットの拡充や、他の法律に対応付けにも効果を発揮するかを調査する必要がある。

4.3 日本法と外国法の対応付け

表 2 に各モデルの評価を示す。MiniML と XLM-r が BERT を僅かに超える結果となった。LaBSE が BERT より低い結果となったが、これは LaBSE の対応言語数の多さによるものだと考えられる。結果から、多言語モデルを用いることで単言語に統一して行っていた外国法との対応付けを、法律原文のまま対応付けることが可能であることが示唆された。

表 1 各手法の対応付け性能評価

Method	Accuracy	Precision	Recall	F1-score
Jaccard	0.870	0.871	0.812	0.840
BM25	0.841	0.856	0.754	0.802
BERT-pretrained	0.838	0.867	0.731	0.794
SBERT-TSDAE	0.875	0.877	0.817	0.846
SBERT-STS	0.895	0.911	0.836	0.872

表 2 日本民法からみたドイツ民法との対応付け結果

Model	Correct Answers
bert-base-uncased	33 / 106 = 0.311
LaBSE	30 / 106 = 0.283
paraphrase-multilingual-MiniML-L12-v2	34 / 106 = 0.321
paraphrase-xlm-r-multilingual-v1	38 / 106 = 0.358

5 おわりに

本研究では法律間の類似条文の対応付けにおいて、fine-tuning の有効性を検証した。実験より、教師あり学習を行ったモデルは従来の BERT の評価を大きく上回り、最も高い性能を発揮した。教師なし学習である TSDAE についても、法令テキストのみで性能を向上させることができるため有効であることを確認した。今後は多言語モデルの fine-tuning を行い、外国法との対応付けでも性能向上を目指せるかの検証を行っていきたい。

謝辞

本研究は、(財)内田エネルギー科学振興財団の助成(助成番号 R05-1008)を受けた。

参考文献

- [1] Devlin, J. et al. (2019). "BERT: Pre-training of deep bidirectional transformers for language understanding." In Proceedings of NAACL, pp. 4171–4186.
- [2] 滝沢 正, 比較法, 三省堂 (2020)
- [3] 長裕樹, 中村誠 (2021). 比較法研究における外国法との類似条項の対応付けと翻訳精度との関係について. 2021 年度電子情報通信学会信越支部大会, p. 115.
- [4] 長裕樹, 中村誠 (2022). BERT を用いた比較法研究における類似条項の対応付け. 言語処理学会第 28 回年次大会発表論文集, pp. 948-951.

- [5] 小関龍也, 長裕樹, 中村誠 (2023). Doc2Vec と BERT を用いた比較法研究における類似条項の対応付け. 言語処理学会第 29 回年次大会発表論文集, pp. 182-186.
- [6] Lee, J., et al. (2020). "BioBERT: a pre-trained biomedical language representation model for biomedical text mining." *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240.
- [7] Reimers, N., et al. (2019). "Sentence-BERT: Sentence embeddings using Siamese BERT-networks." In *EMNLP-IJCNLP*, pp. 3982-3992.
- [8] Wang, K., et al. (2021). "TSDAE: Using Transformer-based Sequential Denoising Auto-Encoder for Unsupervised Sentence Embedding Learning." In *EMNLP2021*, pp. 671–688.
- [9] Vaswani, A., et al. (2017). "Attention is all you need." In *Advances in Neural Information Processing Systems*, pp. 6000–6010.
- [10] 二宮周平(編) (2017). 新注釈民法(17)親族(1) 725 条～791 条. 有斐閣.
- [11] 中川善之助, 山嶋正男(編) (1994). 新版注釈民法(24) 親族(4) 792 条～817 条の 11. 有斐閣.
- [12] 於保不二雄, 中川淳(編) (1994). 新版注釈民法(25)親族(5) 818 条～881 条. 有斐閣.
- [13] Cer et al. (2017). SemEval-2017 Task 1: Semantic Textual Similarity Multilingual and Crosslingual Focused Evaluation. In *SemEval-2017*, pp. 1-14.