

法令間委任関係におけるマルチラベル委任先条文特定

藤田剛 澤田悠冶 坂井優介 渡辺太郎

奈良先端科学技術大学院大学

fujita.tsuyoshi.fy4@naist.ac.jp

{yuya.sawada.sr7, sakai.yusuke.sr9, taro}@is.naist.jp

概要

日本の法令では、形式的効力が上位の法令が下位の法令に対して詳細事項の規定を委任する方式が広く用いられている。とりわけ、一つの条文から複数の下位法令・条文へ委任が行われる場合、読解時には多くの関連条文を横断的に参照する必要があり、読み手の負担は大きい。本稿では、委任関係を含む法令の読解を支援するシステムの開発に向け、委任先が複数存在する状況に焦点を当てて、委任先候補検索モデルと、各候補の適否を判定する二値分類モデルを組み合わせ、全委任先の自動特定を試みる。実験の結果、提案手法により複数の委任先を細かい粒度で特定する性能が向上することを確認した。

1 はじめに

日本の法令には、形式的効力が上位の法令が下位の法令に対して詳細事項の規定を委ねる「委任関係」が多く存在する(図1)。委任関係の把握は法制度の理解に不可欠である一方、委任関係を含む法令の読解では、読み手が委任先法令を頻繁に参照する必要があり、読み手の負担が大きい。そのため、条文中の委任関係を読み手に対して即時に提示する支援システムが求められている。しかし、支援システムの開発・運用のためには、新規法令の制定や既存法令の改正の度に全法令に対する委任関係のアノテーションが必要であり、長時間の作業を要する。

また、委任関係においては、委任先条文と委任元条文との間に一対一の対応が存在する場合もあれば、一つの委任元条文が複数の条文に対して同時に委任を行う場合も存在する。例えば、図1¹⁾で示すように、上位法令である「外国為替及び外国貿易法」の第68条の2で、「政令の定めるところにより、... 権限の一部を地方支分部局の長に委任することが

1) 本稿では、条文の字下げ等の配字に関する詳細を簡略化している。

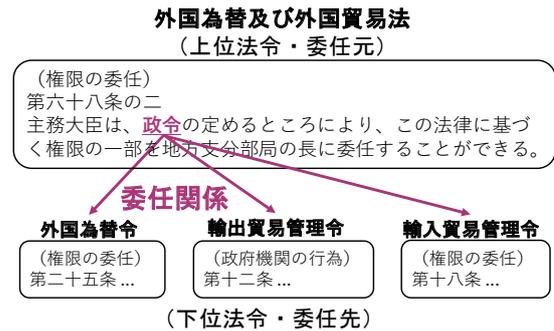


図1 委任関係の例

できる」という文言が出現し、下位法令である「外国為替令」、「輸出貿易管理令」及び「輸入貿易管理令」という三つの政令の条文で、委任される権限の詳細が定められるといった形である。委任先条文が複数存在する場合、委任関係のアノテーションにおいて、多くの法令の中に点在する条文を横断的に確認する必要があり、委任先条文が一つの場合よりもアノテーションにかかる負荷が格段に大きい。

藤田ら[1]は、委任関係を表すキーワードを抽出する固有表現抽出モデル[2, 3]と、各キーワードが示す委任先条文の特定を行う類似度ベースのエンティティ曖昧性解消モデル[4, 5, 6, 7]のパイプライン処理[8, 9, 10, 11]により委任関係の自動抽出を行った。藤田らは、委任関係キーワードの抽出で95ポイント以上のF値、委任先条文の特定で90ポイント以上のRecall@10を報告しており、人手アノテーション支援の用途においては十分な精度を達成している。ただ、藤田らの研究では、委任先が複数存在する場合でもいずれか一つをモデルが予測できれば正解とみなす“シングルラベル設定”で実験が行われており、全ての委任先を予測する“マルチラベル設定”への対応は課題として残っていた。

そこで本稿では、マルチラベル委任先条文特定という新しいタスクを提案し、全委任先の特定を行うことで、アノテーション負荷が特に高いマルチラベ

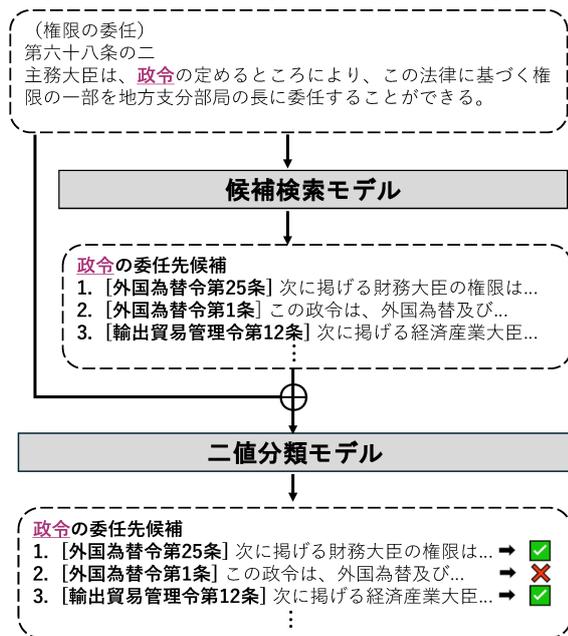


図2 マルチラベル委任先条文特定における処理

ル事例に対応し、より包括的なアノテーション支援の実現を目指す。具体的には、委任関係キーワードと委任先候補の埋め込み表現の類似度により、複数の委任先を考慮して委任先候補の絞り込みを行う候補検索モデルと、絞り込まれた後の各委任先候補が委任先として適切かどうかを判断する二値分類モデルを繋ぎ合わせたパイプライン処理を行う。実験の結果、パイプライン処理により、候補検索モデル単体と比較して、項や号といった細かい粒度で全委任先を特定する性能が向上することが確認できた。

2 手法

2.1 タスク設定

マルチラベル委任先条文特定タスクは、上位法令の条文及び条文中の委任関係キーワードの位置を入力とし、キーワードが示す委任先の下位法令の条文を全て特定する。例えば、図1の事例では、入力として外国為替及び外国貿易法第68条の2の条文と、同条文中の委任関係キーワード「政令」の位置が入力となり、委任先条文として外国為替令第25条、輸出貿易管理令第12条及び輸入貿易管理令第18条を特定する。本稿は、図2に示すように、委任先候補の絞り込みを行う候補検索モデルと、各候補が委任先として適切かを予測する二値分類モデルを組み合わせたパイプライン処理により、マルチラベル設定での委任先条文の特定を行う。

2.2 候補検索

本稿では、藤田ら [1] の委任先候補検索モデルをマルチラベルに対応するよう拡張し、委任元条文中の委任関係キーワードが示す委任先条文の候補を検索する。藤田らの委任先候補検索モデルは、条文中の委任関係キーワードの位置が埋め込まれた条文 x と、委任先候補条文の説明文 y を Text Embedding モデル [12, 13] により埋め込み表現 h_x 及び h_y に変換し、 $s(x, y) = h_x^T h_y$ により計算される内積を x と y の類似度として扱うことで、上位の類似度を示す条文を委任先候補として抽出する。モデル学習においては、式1の損失関数を使用し、一つの正解委任先条文に対して類似度を最大化する。

$$\mathcal{L}_{\text{single}} = -\log \left(\frac{\exp(s(x, y^+))}{\sum_{y' \in Y} \exp(s(x, y'))} \right) \quad (1)$$

ここで y^+ は正解の委任先候補条文を表し、正解が複数存在する場合はそのうち一つを正解条文 y^+ とみなしている。また、 y' は委任先候補集合 Y に含まれる候補条文 ($y' \in Y$) を表し、 Y はミニバッチ内の候補条文集合 Y_C と y^+ に対する Hard-Negative サンプル集合 Y_{hard} からなる ($Y = Y_C \cup Y_{\text{hard}}$) [7, 14]。

本稿は、複数の委任先条文を考慮したモデル学習を行うため、式2及び式3の損失関数 $\mathcal{L}_{\text{multi}1}$, $\mathcal{L}_{\text{multi}2}$ を導入して性能を比較する。

$$\mathcal{L}_{\text{multi}1} = -\log \left(\frac{\sum_{y^+ \in Y_P} \exp(s(x, y^+))}{\sum_{y' \in Y_P \cup Y_N} \exp(s(x, y'))} \right) \quad (2)$$

$$\mathcal{L}_{\text{multi}2} = -\sum_{y^+ \in Y_P} \log \left(\frac{\exp(s(x, y^+))}{\sum_{y' \in Y_N \cup \{y^+\}} \exp(s(x, y'))} \right) \quad (3)$$

ここで Y_P と Y_N はそれぞれ、委任先候補条文集合 Y 中の候補条文のうち、正解の候補条文及び不正解の候補条文の集合 ($Y = Y_P \cup Y_N$) である。 $\mathcal{L}_{\text{multi}1}$ は、シングルラベル設定の式1の拡張であり、 Y_P に含まれる全ての条文のスコアを総和し、正解の条文全体が不正解の条文よりも高い類似度を持つようにモデル学習を行う。 $\mathcal{L}_{\text{multi}2}$ は、Zhang ら [15] のマルチラベル候補検索モデルの目的関数であり、各正解条文を独立に扱い、各正解条文が不正解条文よりも高い類似度を持つようにモデルを学習する。

2.3 二値分類

二値分類モデルは、Cross Encoder [7] を二値分類タスク向けに変換し、委任元条文と委任先候補条文の

ペアの間に委任関係が存在するか否かを判定する。具体的には、エンコーダ型言語モデル (Encoder) と全結合層 (Liner) と連結し、候補条文 y が委任関係キーワード x の委任先である確率を式 4 及び式 5 で計算する。

$$P(y | x) = \sigma(z) \quad (4)$$

$$z = \text{Liner}(\text{cls}(\text{Encoder}(\tau_{x,y}))) \quad (5)$$

ここで、 $\tau_{x,y}$ は Wu ら [7] と同様の形式で x と y を結合したトークン系列、 $\text{cls}(\cdot)$ は [CLS] トークンに対するエンコーダの出力を取得する関数、 $\sigma(\cdot)$ はシグモイド関数である。

学習時には、各正解候補条文に対する類似度が最大となるように、二値交差エントロピー誤差 (式 6) の最小化によってパラメータを更新する。

$$\mathcal{L}_{\text{BCE}} = -\left[\mathbb{1}(y) \log P(y | x) + (1 - \mathbb{1}(y)) \log(1 - P(y | x)) \right] \quad (6)$$

ここで、 y が正解の委任先条文である場合は $\mathbb{1}(y) = 1$ 、それ以外の場合は $\mathbb{1}(y) = 0$ である。推論時には、 $P(y | x)$ が閾値を上回った正解候補条文を x の委任先として推定する。

3 データセット

本稿では、藤田ら [1] が作成した法令条文データベース及び委任関係抽出データセットを使用して、マルチラベル委任先条文特定モデルの学習と評価を行う。法令条文データベースには、日本における現行法令²⁾を階層的に構成する様々な粒度の条文のうち、法令全体・条・項・号・附則全体の粒度を持つもの約 228 万件が収録され、各条文に固有の条文 ID が設定されている。また、委任関係抽出データセットは、全現行法令の条文を句点で区切って得られる文のうち、委任関係キーワードを含むものの一部である 20,386 文から構成されている。各文には、文中に出現する委任関係キーワード計 20,723 件の位置情報と、各キーワードに対応する 35,035 件の委任先条文 ID が正解ラベルとして人手でアノテーションされている³⁾。

委任関係抽出データセットにおける、各委任関係キーワードに対応する正解の委任先条文数の分布を表 1 に示す。データセット中のキーワードのうち複数の委任先条文を持つものは約 14% 程度であり、大

2) 国会が制定する法規範である法律と、国の行政機関が制定する法規範である命令からなる。

3) 法令における条文の粒度及び本稿で使用するデータセットに関する詳細は付録 A を参照。

表 1 委任先条文数

正解の委任先条文数	1	2-10	11-100	101-
委任関係キーワード数	17,890	2,508	320	5
割合 (%)	86.3%	12.1%	1.54%	0.02%

半は委任先条文数が 2-10 個の範囲に集中している。委任先条文が複数存在する事例の割合は比較的低いものの、これらの事例は、委任関係の人手アノテーションにおいて特に大きな負担を生じさせる (1 節参照)。そのため、複数の委任先を持つ事例への対応は実務上重要な課題である。なおマルチラベル委任先条文特定タスクでは、法令条文データベース中の条文を委任先候補とし、委任関係抽出データセットに含まれる各キーワードの正解ラベルを委任先候補の中から全て予測する。

4 実験設定

本稿では、藤田ら [1] と同様に、委任関係抽出データセットを用いて「法令全体」、「条」、「項」及び「号」の 4 つの粒度でスコアを計算し、5 分割交差検証によって候補検索モデルと二値分類モデルの性能を評価する。5 分割交差検証では、全体の 2 割の文を評価データとして使用し、残りの 8 割のデータをさらに 9:1 の割合で訓練データと開発データに分割して使用する⁴⁾。

候補検索 多言語版 E5-Base モデル⁵⁾ [13] をエンコーダとして使用し、シングルラベル設定の損失関数 $\mathcal{L}_{\text{single}}$ (式 1)、マルチラベル設定の損失関数 $\mathcal{L}_{\text{multi1}}$ (式 2) 及び $\mathcal{L}_{\text{multi2}}$ (式 3) により学習した候補検索モデル Ret-Single, Ret-Multi1 及び Ret-Multi2 の性能を比較する。表 1 のように、各委任関係キーワードに対する委任先条文数は大半が 10 件以下であるため、推論時に候補検索モデルが出力する候補の件数は、正解の委任先条文全てを十分に包含できる 30 件とする。評価においては、類似度上位 30 件の候補に対して Recall@30 (R@30) を算出し、正解の委任先条文のうち、上位 30 件の候補に含まれる正解条文の割合を比較する。

二値分類 二値分類モデルでは、エンコーダ型言語モデルとして東北大学版の日本語 BERT⁶⁾ を使用し、Ret-Single, Ret-Multi1 及び Ret-Multi2 によって検索された各 30 件の委任先候補条文から二値分類モデル Bin-Single, Bin-Multi1 及び Bin-Multi2 をそれ

4) 学習及び評価における詳細設定は付録 B に示す。

5) <https://huggingface.co/intfloat/multilingual-e5-base>

6) [tohoku-nlp/bert-large-japanese-v2](https://github.com/tohoku-nlp/bert-large-japanese-v2)

表2 候補検索モデルのR@30

粒度	Ret-Single	Ret-Multi1	Ret-Multi2
法	95.2 (±0.91)	91.8 (±1.13)	91.8 (±1.00)
条	72.2 (±2.16)	63.6 (±2.26)	65.4 (±1.31)
項	69.5 (±2.43)	60.0 (±2.28)	62.0 (±1.13)
号	68.5 (±2.58)	59.3 (±2.33)	61.1 (±1.24)

ぞれ作成する。またベースラインとして、Ret-Singleが検索した候補の中で最も類似度が高かった条文を直接推論結果として出力するBase-Singleも比較対象に加える。評価実験では、モデルが予測した委任先条文と正解の委任先条文との間の一致について、適合率、再現率及びF値を算出する。

パイプラインシステム パイプラインシステムでは、Ret-SingleとBin-Single, Ret-Multi1とBin-Multi1, Ret-Multi2とBin-Multi2をそれぞれ組み合わせた、Pipe-Single, Pipe-Multi1及びPipe-Multi2による推論を行い、正解の委任先条文との間の一致について、適合率、再現率及びF値を算出する。

5 実験結果と考察

候補検索 候補検索モデルの評価実験の結果は表2の通りである。全ての評価粒度において、シングルラベル設定の損失関数 $\mathcal{L}_{\text{single}}$ により学習したRet-Singleが、マルチラベル設定を明示的に考慮したRet-Multi1及びRet-Multi2を上回るR@30を示した。特定の委任元条文に対応する正解条文は全て、その委任元条文に関連する詳細事項を規定するという点で共通であるため、互いの意味的な類似性が高い場合が多い。そのため、 $\mathcal{L}_{\text{single}}$ により複数の正解条文から一つを選択して学習に使用した場合でも、選択されなかった条文も含めた全ての正解条文に対して、モデルが高い類似度スコアを割り当てられるようになった可能性がある。

二値分類・パイプラインシステム パイプラインシステムの性能を表3に示す。パイプラインシステムは、「項」及び「号」の粒度においてBase-Singleを上回っており、二値分類モデルの導入により、複数の委任先を項や号といったより細かい粒度で特定する能力が向上したことが分かる。また、Pipe-Singleは、Ret-Singleの出力する候補条文の中により多くの正解条文が含まれているため、Pipe-Multi1及びPipe-Multi2よりも高い再現率を示している。他方、適合率はPipe-SingleよりもPipe-Multi1及びPipe-Multi2の方が高く、結果として3種類のパイプラインシステムのF値は同程度の値になった。

表3 パイプラインシステムの性能

モデル	粒度	適合率	再現率	F値
Base-Single	法	92.6 (±0.89)	85.3 (±1.15)	88.8 (±0.93)
	条	78.9 (±1.84)	47.3 (±2.52)	59.1 (±2.35)
	項	56.3 (±6.64)	33.3 (±3.40)	41.8 (±4.45)
	号	56.0 (±6.70)	33.1 (±3.44)	41.6 (±4.50)
Pipe-Single	法	83.1 (±3.58)	78.1 (±4.57)	80.5 (±3.97)
	条	64.3 (±9.67)	53.2 (±4.70)	57.9 (±5.66)
	項	56.2 (±10.5)	48.9 (±5.61)	51.8 (±6.61)
	号	56.2 (±10.5)	48.9 (±5.61)	51.7 (±6.62)
Pipe-Multi1	法	87.0 (±4.06)	72.9 (±8.32)	79.2 (±6.58)
	条	71.6 (±6.27)	48.6 (±8.05)	57.8 (±7.77)
	項	60.8 (±10.2)	45.3 (±8.09)	51.9 (±9.00)
	号	60.7 (±10.2)	45.3 (±8.09)	51.9 (±9.01)
Pipe-Multi2	法	85.4 (±7.00)	75.8 (±4.71)	80.0 (±3.72)
	条	70.1 (±9.54)	51.1 (±5.01)	58.6 (±5.10)
	項	62.4 (±8.76)	47.6 (±4.65)	53.6 (±4.76)
	号	62.4 (±8.76)	47.6 (±4.65)	53.6 (±4.76)

Ret-Singleは、複数の正解条文の中で一つの条文のみを正解として扱い、その条文と類似度が高い候補条文を検索するように学習している。そのため、Ret-Singleが出力する候補条文には、特定の正解条文との意味的類似性がマルチラベル設定と比べて高く、正解との識別が比較的難しい不正解条文も多く含まれている可能性がある。そして、Bin-Singleがこれらの比較的識別難易度が高い候補条文に対応できず、不正解条文を正解の委任先として誤検出する事例が増加したことで、Pipe-Singleの適合率の低下につながった可能性が考えられる。候補検索モデルのR@30に差がある一方でパイプラインシステムのF値が同程度となることは、正解との意味的類似性が高い不正解条文に対する二値分類モデルの識別性能向上が、候補検索モデルの出力の網羅性を適合率の向上に繋げ、パイプライン全体の性能改善を図るために重要であることを示唆している。

6 おわりに

本稿は、委任関係の人手アノテーションにおいて特に負荷が高いマルチラベル事例に対処するため、委任先条文を全て抽出するマルチラベル委任先条文特定タスクを提案した。そしてこのタスクに取り組むため、候補検索モデルと二値分類モデルを組み合わせたパイプラインシステムを構築した。評価実験の結果、パイプラインシステムにより項や号といった細かい粒度における委任先特定性能が向上することが示され、委任関係アノテーション支援における提案手法の有効性が示唆された。

謝辞

本研究は、第一法規株式会社と奈良先端科学技術大学院大学の共同プロジェクトとして実施したものです。

参考文献

- [1] 藤田剛, 澤田悠治, 坂井優介, 渡辺太郎. 法令文における法令間委任関係の自動抽出. 研究報告自然言語処理 (NL), 2025-NL-266(25), pp. 1–18, 2025. <https://ipsj.ixsq.nii.ac.jp/records/2006080>.
- [2] Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, and Yuji Matsumoto. LUKE: Deep contextualized entity representations with entity-aware self-attention. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 6442–6454, Online, November 2020. Association for Computational Linguistics.
- [3] 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 Tamar 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.
- [4] Riccardo Orlando, Pere-Lluís Hugué Cabot, Edoardo Barba, and Roberto Navigli. ReLiK: Retrieve and LinK, fast and accurate entity linking and relation extraction on an academic budget. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 14114–14132, Bangkok, Thailand, August 2024. Association for Computational Linguistics.
- [5] Hibiki Nakatani, Hiroki Teranishi, Shohei Higashiyama, Yuya Sawada, Hiroki Ouchi, and Taro Watanabe. A text embedding model with contrastive example mining for point-of-interest geocoding. In Owen Rambow, Leo Wanner, Marianna Apidianaki, Hend Al-Khalifa, Barbara Di Eugenio, and Steven Schockaert, editors, *Proceedings of the 31st International Conference on Computational Linguistics*, pp. 7279–7291, Abu Dhabi, UAE, January 2025. Association for Computational Linguistics.
- [6] Daniel Gillick, Alessandro Presta, and Gaurav Singh Tomar. End-to-end retrieval in continuous space, 2018.
- [7] Ledell Wu, Fabio Petroni, Martin Josifoski, Sebastian Riedel, and Luke Zettlemoyer. Scalable zero-shot entity linking with dense entity retrieval. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 6397–6407, Online, November 2020. Association for Computational Linguistics.
- [8] Riccardo Pozzi, Riccardo Rubini, Christian Bernasconi, and Matteo Palmonari. Named entity recognition and linking for entity extraction from italian civil judgements. In Roberto Basili, Domenico Lembo, Carla Limongelli, and Andrea Orlandini, editors, *AIXIA 2023 – Advances in Artificial Intelligence*, pp. 187–201, Cham, 2023. Springer Nature Switzerland.
- [9] Manoj Prabhakar Kannan Ravi, Kuldeep Singh, Isaiah Onando Mulang^{*}, Saeedeh Shekarpour, Johannes Hofbart, and Jens Lehmann. CHOLAN: A modular approach for neural entity linking on Wikipedia and Wikidata. In Paola Merlo, Jorg Tiedemann, and Reut Tsarfaty, editors, *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pp. 504–514, Online, April 2021. Association for Computational Linguistics.
- [10] Johannes M. van Hulst, Faegheh Hasibi, Koen Dercksen, Krisztian Balog, and Arjen P. de Vries. Rel: An entity linker standing on the shoulders of giants. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '20*, p. 2197–2200, New York, NY, USA, 2020. Association for Computing Machinery.
- [11] 澤田悠治, 安井雄一郎, 大内啓樹, 渡辺太郎, 石井昌之, 石原祥太郎, 山田剛, 進藤裕之. 企業名の類似度に基づく日経企業 id リンキングシステムの構築と分析. 自然言語処理, Vol. 31, No. 3, pp. 1330–1355, 2024.
- [12] Liang Wang, Nan Yang, Xiaolong Huang, Binxing Jiao, Linjun Yang, Daxin Jiang, Rangan Majumder, and Furu Wei. Text embeddings by weakly-supervised contrastive pre-training, 2024.
- [13] Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. Multilingual e5 text embeddings: A technical report, 2024.
- [14] Tsuyoshi Fujita, Yuya Sawada, Yusuke Sakai, and Taro Watanabe. Toward automatic delegation extraction in japanese law. In *EACL 2026 Industry Track*. Association for Computational Linguistics, March 2026. (To Appear).
- [15] Wenzheng Zhang, Wenyue Hua, and Karl Stratos. EntQA: Entity linking as question answering. In *International Conference on Learning Representations*, 2022.

A 日本の法令における条文の粒度

日本の法令は、図3（文献[1]より引用）のように「条」・「項」・「号」といった粒度の異なる要素が階層構造を成すものと見ることが可能である。この階層構造においては、上位の要素が下位の要素を複数包含しうる。例えば、ある法令の「第一条」の中には「第一項」や「第二項」等が存在し、「第二項」の中には「第一号」や「第二号」等が存在するといった形である。本稿で使用する委任関係抽出データセットに含まれる正解の委任先条文ラベルと、法令条文データベース中の条文の粒度別内訳は、表4（文献[1]より引用）の通りである。

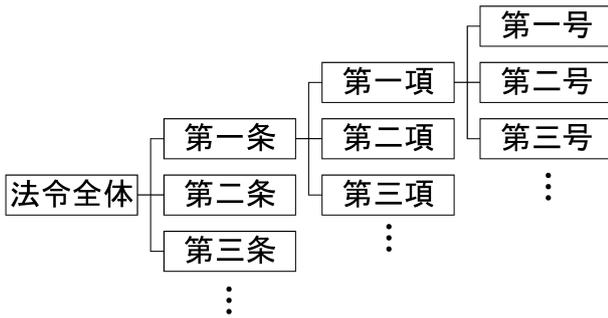


図3 法令中の要素間の階層構造

表4 委任先条文ラベルと条文データベースの粒度別内訳。委任先条文ラベルは延べ個数を記す。

粒度	委任先条文ラベル	条文データベース
法令全体	5,169	29,788
条	16,923	418,855
項	12,906	1,020,253
号	25	599,426
附則全体	12	208,252
合計	35,035	2,276,574

B 実験設定の詳細

B.1 候補検索モデル

候補検索モデルの学習時に使用したハイパーパラメータは表5の通りである。また、候補検索モデルの評価指

表5 候補検索モデル学習時のハイパーパラメータ

	Ret-Multi1	Ret-Multi2	Ret-Single
batch size (train)	4		16
batch size (eval)	16		16
learning rate	1e-5		1e-5
epochs	1		4
optimizer	AdamW		AdamW
eps	1e-6		1e-6
scheduler	linear		linear
warmup ratio	0.06		0.06
weight decay	0.01		0.01
max grad norm	0.00		0.00
β	[0.9, 0.98]		[0.9, 0.98]

表6 二値分類モデル学習時のハイパーパラメータ

batch size (train)	2
batch size (eval)	8
learning rate	5e-5
epochs	4
optimizer	AdamW
eps	1e-6
scheduler	linear
warmup ratio	0.1
weight decay	0.01
max grad norm	0.00
β	[0.9, 0.98]

標として使用する R@30 は式7で定義される。

$$R@30 = \frac{\sum_{i=1}^N |G_i \cap C_i^{30}|}{\sum_{i=1}^N |G_i|} \quad (7)$$

ここで、 i は委任関係キーワードを表すインデックスであり、 N は評価用データ中の委任関係キーワード数、 G_i はキーワード i に対する正解の委任先条文集合、 C_i^{30} はキーワード i に対してモデルが検索した類似度上位 30 件の委任先候補条文集合である。

B.2 二値分類モデル

二値分類モデルの学習時に使用したハイパーパラメータは表6の通りである。また、推論時の二値分類の閾値は 0.5 とし、スコアが 0.5 以上の候補条文を委任先として推定する。モデルの評価指標としては、式8及び式9により計算する適合率及び再現率、そしてこれらの調和平均である F 値を使用する。

$$\text{適合率} = \frac{\sum_{i=1}^N |P_i \cap G_i|}{\sum_{i=1}^N |P_i|} \quad (8)$$

$$\text{再現率} = \frac{\sum_{i=1}^N |P_i \cap G_i|}{\sum_{i=1}^N |G_i|} \quad (9)$$

ここで、 P_i は委任関係キーワード i に対してモデルが予測した委任先候補条文の集合である。

B.3 条文の粒度毎の評価

4節で述べたように、本研究では、藤田ら[1]にならい「法令全体」、「条」、「項」及び「号」の4つの粒度でモデル性能を測る。各粒度における評価では、当該粒度よりも細かい粒度の条文が委任先として指定されている場合、予測結果と正解が当該粒度まで一致していれば正解と判定する。例えば、条レベルの評価において、委任先が項や号で指定されている場合であっても、予測した委任先が正解と同一の条に属していれば正解とみなす。一方で、評価対象の粒度よりも粗い粒度の条文が委任先として指定されている場合には、予測結果が正解と完全に一致する場合にのみ正解と判定する。例えば、条レベルの評価において「法令全体」が委任先である場合、「法令全体」を予測した場合のみを正解とし、当該法令の特定の条を予測した場合は不正解とする。