

法令知識グラフ構築のための法令文における参照関係抽出手法

北野尚樹¹ 天笠俊之²

¹ 筑波大学情報学群情報科学類 ² 筑波大学計算科学研究センター

knaoki@kde.cs.tsukuba.ac.jp

amagasa@cs.tsukuba.ac.jp

概要

法令文中には関係する法令への参照が頻繁に含まれる(例:「第七条の二第一項の場合において、」)。これらの参照情報から、法令をノードとし参照をエッジとする法令知識グラフを作成することができる。しかし、法令文中に含まれる参照は文中で定義された略称に基づくものや「同法第五条」のように直前の参照を援用するものなど、複雑なルールが混在しており簡単に抽出することはできない。

本研究では法令の名称と中身の一覧をもとに、全文検索とパターンベースの解析手法を組み合わせることで、条項の番号を含めて参照を抽出する手法を開発した。また、実験により本手法で参照を高い精度で抽出できることを確認した。

1 はじめに

1.1 法令文中の参照が複雑である背景

法令文中には法令や条項を参照する表現が頻繁に用いられる [1]。参照は他の法令で定められている規則や用語などを援用するために用いられる。援用する規則や用語などが具体的に定義されている場合は条項を指すときには条項番号まで記述して参照をし、法令全体で定義され使われているものである場合には法令名だけを指すようにして参照される。

法令への参照は多くの場合法令の正式な名称と条項番号の組で表現される(図 1)が、中には法令文中で定められた略称が用いられるもの(図 2)、あるいは複数の参照を範囲で指定するものなどの複雑なもの(図 6)など様々なバリエーションがある。このような複雑な参照表現は法令文の複雑化を招き、参照情報の機械的な抽出を困難にさせている。

法令文中の複雑な参照を、略称や条項番号を含めて抽出することができれば、法令知識グラフの作成と法令実務全般の効率化が期待される。

1.2 関連研究

2007 年ごろに、法令文を論理式に変換する過程で参照を抽出する手法が提案されている。「X に規定する Y」や「X の規定により Y」などの文全体のパターンと、X としてありうる「第 a 条第 b 項第 c 号」や「前条第 b 項」などのパターンを組み合わせる検索する手法である [2, 3, 4, 5]。しかし、この手法では条項番号だけは抽出できるものの、法令の名称まで取得することはできておらず、パターンの種類も実際の多様な法令文に対応できていないため、抽出は不完全である。

参照を完全にするためには法令の名称の検索が必要不可欠である。法令名は正式な名称や法令番号だけではなく、法令文中で法令に対して付けられる略称が用いられることがある。略称の定義箇所の検索と抽出は中村らや北野らによって取り組まれている [6, 7, 8]。これらの手法は事前に用意した文型のパターンに従って処理し、正式名称と略称を抽出するものである。これらの手法では文型のパターンに合致しない表現の抽出に失敗することが課題となっている。

略称の定義箇所の検索対象を法令名のものに限定し、法令名の全文検索の結果と組み合わせることで、法令文中に含まれる法令名を高い精度で抽出できる手法を我々が提案した [9]。

近年、大規模言語モデル (LLM) が自然言語処理の分野でも高い性能を示しており、法令文の解析においても大きな可能性があると考えられている [10, 11, 12, 13]。しかし、北野らの報告によると LLM は法令文からテキスト長が長い語を正確に抽出することを不得意としており [8]、テキスト長が長くなりがちな参照を抽出する本タスクには不適切であると考え、我々のパターンベースの法令名抽出手法 [9] を拡張することにした。

1.3 貢献

法令において参照は必要不可欠で頻繁に行われるものであるが、略称や直前の参照の援用などの複雑なものも多くあり、法令実務者にも機械的な抽出においても難しいものであった。本研究では法令文に含まれる参照に対し、網羅的なパターンベースの抽出手法を提案し、その有効性を確認した。本研究の手法により、法令知識グラフの作成や法令実務の効率化などが期待される。

本研究の主な貢献は以下のとおりである。

- 従来の研究では対応できていなかった、法令文に含まれる参照特有の文型を整理し、それらに対応するパターンベースの手法を開発した。
- 本提案手法の F 値 0.9 を超える高精度であることを実験によって確認した。
- 文型の整理と抽出手法の開発の過程で施行中の法令に含まれる脱字を発見し、所管省庁への報告を行なった。

2 諸定義および問題設定

2.1 諸定義

法令の基本的な構成要素は条である。条には法令の中で一意の条番号が振られている。改正により既存の条の間に新しく条を追加するときは、他の条番号を変更させないように、「第〇条の二」のような枝番号と呼ばれる番号が付けられる。条の中の段落を項とよび、それぞれに項番号が振られている。

条数が多い法令には、条文のくくりとして章立てが置かれる場合がある。章立ては、編・章・節・款・目の 5 つの階層からなる。章立てにもそれぞれの章節ごとに「第二章第一節」のような章節番号が与えられる。章節番号にも枝番号が存在する。

本研究で扱う他法令への参照とは、法令文中に含まれる法令の名称と章番号・条項番号のことを指す。法令名だけの場合と、章節番号だけ、条項番号だけの場合もありうる。

例えば図 1 の法令文では船舶安全法という法令名と条項番号で参照を行なっている。また、図 2 の法令文では文中で公認会計士法に「法」という略称を与え、条項番号と併せて参照している。

舟車とは、船舶安全法第二条第一項の規定を適用しない船舶、端舟、はしけ、被曳船その他の舟及び車両をいう。

図 1 法令名と条項番号による参照の例（網掛け部分が参照）

公認会計士の登録を受けようとする者は、公認会計士法（昭和二十三年法律第百三三号。以下「法」という。）第十条に規定する実務補習の外に、法第二条第一項の業務について公認会計士（外国公認会計士及び外国において公認会計士の資格に相当する資格を有する者を含む。以下同じ。）若しくは監査法人を補助すること（中略）を必要とする。

図 2 法令文中で定義した略称と条項番号を使った参照（網掛け部分が参照）

2.2 問題設定

法令が N 件あるとする。このとき、法令全体の集合を $L = \{l_1, l_2, \dots, l_N\}$ と表す。 l_i で参照されている他の法令への参照の集合を

$$P_i = \{(l_j, v) \mid l_j \in L, j \neq i, v \text{ は法令中の位置}\}$$

とする。

本研究で取り組む問題は、 l_i のテキストから P_i を生成することである。

3 法令文の観察

我々による法令名を抽出する手法 [9] では法令名の複雑な出現パターンが整理された。本研究では、それに加えて章節番号や条項番号の出現パターンを観察し次のように整理した。

3.1 前後の条項への参照

前後の条項への参照は次の語が用いられる。

- 前条
- 前 n 条
- 前各条
- 前項
- 前 n 項
- 前各項
- 次条
- 次項

直前から n 個の条項をまとめて参照する場合は「前 n 条」や「前 n 項」が用いられる。また、第一条からの全てをまとめて参照する場合は前各条、第一

項からの全てをまとめて参照する場合は前各項を用いる。直後の条項を参照する場合は次条や次項を用いる。

3.2 直前の参照の援用

直前に参照された法令や条項を再引用し、場合によっては援用して新たに参照を作成することがある(図3)。それに用いられるのが次の語である。

- 同法
- 同令
- 同条
- 同項

森林法(明治四十年法律第四十三号)(以下「旧法」という。)第九条(中略)及び同法第十条第一項の規定により

図3 直前の参照の援用(網掛け部分が参照とその援用)

3.3 同じ法令内の他の条項の参照

直前に参照がなく単独で章節番号や条項番号がある場合は、同じ法令内の他の条項を参照しているものである。

〇〇について第三条第三項の規定により

図4 同じ法令内の章節番号や条項番号の参照(網掛け部分が章節番号と条項番号)

3.4 並列

法令名の後に章節番号や条項番号が並列で続くことがある。並列は読点と「又は」・「及び」・「並びに」・「若しくは」の語によって行う(図5)。

この法律の規定は、第二章、第三十五条及び次章の規定を除き、

図5 章節番号や条項番号の並列(網掛け部分が章節番号と条項番号)

3.5 範囲

複数の連続する章節や条項を参照する場合は「〇〇から△△まで」もしくは「〇〇乃至△△」で範囲を指定して行う[1](図6)。

4 提案手法

3節の観察の結果により、章節番号や条項番号は主に法令の出現箇所と「同法」や「同条」などの相対的な参照の直後に出現することがわかった。章節

旧法第九条から第十一条ノ二まで、
民法第二百四十二条乃至第二百四十六条ノ規定ニ

図6 章節番号や条項番号の範囲の指定(網掛け部分が章節番号と条項番号)

番号や条項番号が単独で出現するときは同じ法令内の該当する箇所を指している。

そのため、参照を重複なく抽出するために次の手順を取ることにした。

1. 我々の法令名抽出手法[9]を使って法令の参照箇所を抽出する(図7ステップ1-3)。
2. 3.1節や3.2節の観察に則り、前後の条項への参照や直前の参照を援用する語を検索する(図7ステップ4)。
3. 法令の参照箇所の直後に章節番号や条項番号が出現するかどうかを図5に則って探索する(図7ステップ5)。
4. 単独で出現する章節番号や条項番号を探す(図7ステップ6)。
5. 抽出した法令と章節番号・条項番号を組み合わせる完全な参照にし、前後の条項への参照なども具体的な参照に置き換える(図7ステップ7)。
6. 3.5節で述べたような範囲で行われる参照を全て具体的な参照のリストに直す(図7ステップ8)。

ただし、法令文中には改正や準用のためにカギ括弧でくくることがある[1][14]。このカギ括弧の中のテキストは法令に対する参照ではなく単なる引用の文字列であるため、抽出する参照の対象から外す必要がある[9]。そこで、事前に開きカギ括弧と閉じカギ括弧で囲われている箇所の範囲を確認しておき、前述の全てのステップにおいてこの範囲に当てはまるものを除外する操作を行う。

図7ステップ5やステップ6では枝番号も含めた章節番号や条項番号を正規表現で探索する。このときに用いる正規表現は次のものとなる。

第[一二三四五六七八九十百千]+(編|章|節|款|目|条|項)(((|ノ)[一二三四五六七八九十百千]+)*)

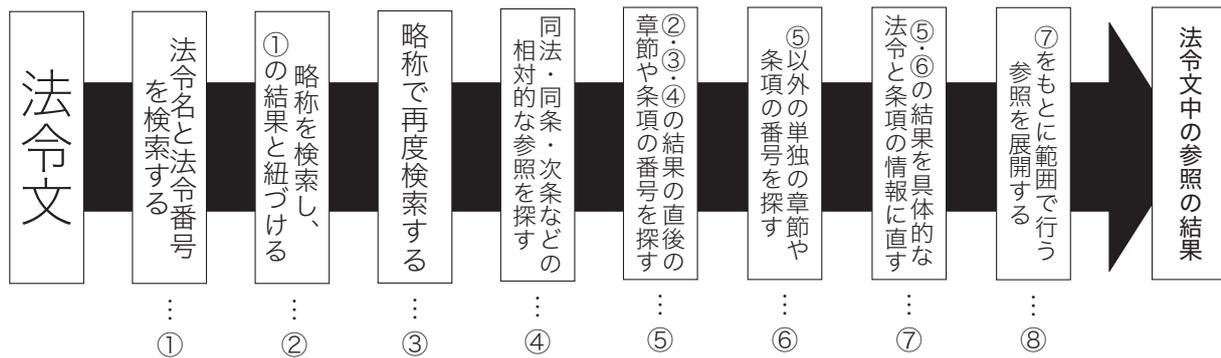


図7 法令文中から参照を抽出する手法の各ステップの詳細

5 実験

5.1 実験方法・内容

使用する法令名・法令番号・法令のテキストは日本政府が法令データを公開している e-Gov 法令検索¹⁾の 2025 年 12 月 20 日時点の法令の XML データを使用する。この法令の XML データから法令を 100 個ランダムに抽出した。これらに対し提案手法を適用し、他法令への参照を抽出した。抽出結果に対し手動で正誤の判定を行った。

5.2 実験結果・考察

ランダムに選んだ 100 件の法令のうち、参照は 6800 個存在していた。このうち、参照先の法令名や条項番号まで含めて正確に抽出することができていたものは 6006 個であった。参照として抽出することはできたものの参照先の法令名や条項番号に誤りがあったものが 721 個、参照ではないが参照として抽出してしまったものが 44 個であった。参照であるにも関わらず抽出できなかったものは 73 個であった。

この値から計算すると式 1, 2, 3, 4 より、適合率が 0.887、再現率が 0.988、F 値が 0.935、法令名・条項番号の紐付けの正確さが 0.893 となった。

再現率が 0.988 であることや参照の抽出の精度が 0.989 であることから、本手法では法令文中の参照をほとんど抽出できることがわかった。一方で、適合率は 0.887 にとどまっており、法令名・条項番号の紐付けの正確さが 0.893 であることから、紐付けの正確さが誤りの要因であることがわかった。

$$\text{適合率} = \frac{6006}{6006 + 721 + 44} \approx 0.887 \quad (1)$$

$$\text{再現率} = \frac{6006}{6006 + 73} \approx 0.988 \quad (2)$$

$$F \text{ 値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}} \approx 0.935 \quad (3)$$

$$\text{法令名・条項番号の紐付け精度} = \frac{6006}{6006 + 721} \approx 0.893 \quad (4)$$

6 結論

本研究では、法令文中の法令名や条項番号などからなる参照を抽出するための手法を提案した。正規表現による全文検索などを駆使するパターンベースの比較的単純で軽量な手法で、F 値が 0.935 という高い精度を実現した。

これにより、法令の条文同士の参照関係をもとにした法令知識グラフの構築が可能となり、検索の精度向上などに大きく貢献できると期待できる。

本研究での参照抽出手法を開発中に、動物用医療機器及び動物用体外診断用医薬品の製造管理及び品質管理に関する省令（平成七年農林水産省令第四十号）第三十二条や国土交通省関係住宅宿泊事業法施行規則（平成二十九年国土交通省令第六十五号）第十六条中の参照の中に誤字脱字を発見した。所管省庁である農林水産省や国土交通省に問い合わせたところ、誤記であることが確認され、次の改正のタイミングで訂正される旨が伝えられた。

このことから、本手法では入力への誤りに対して頑健であり、法制執務に対して実際に寄与できる性能であることが確認された。

1) <https://laws.e-gov.go.jp/bulkdownload/>

参考文献

- [1] 磯崎陽輔. 分かりやすい法律・条例の書き方 [改訂版]. 株式会社ぎょうせい, 2017.
- [2] 江尻暁, 北田安希雄, 島津明. 法令文の論理式への変換-論理構造について-. 言語処理学会 第 12 回年次大会 発表論文集, pp. 624-627, 2006.
- [3] 北田安希雄, 江尻暁, 島津明. 法令文の論理式への変換-原子文について-. 言語処理学会 第 12 回年次大会 発表論文集, pp. 628-631, 2006.
- [4] 信岡俊祐, 中村誠, 島津明. 法令文の論理式への変換. 言語処理学会 第 13 回年次大会 発表論文集, pp. 254-257, 2007.
- [5] 木村祐介, 中村誠, 島津明. 号の列举や参照表現をもつ法令文の論理式への変換. 言語処理学会 第 14 回年次大会 発表論文集, pp. 612-615, 2008.
- [6] Makoto Nakamura, Ryusei Kobayashi, Yasuhiro Ogawa, and Katsuhiko Toyama. A pattern-based approach to hyponymy relation acquisition for the agricultural thesaurus. In **Proceedings of International Symposium on Agricultural Ontology Service 2012 (AOS2012)**, 2012.
- [7] 中村誠, 小川泰弘, 外山勝彦. 法令文中において括弧書きで定義されている法令用語とその語釈文の抽出. 言語処理学会 第 19 回年次大会 発表論文集, 2013.
- [8] 北野尚樹, 西山大輝. 法令文の可読性向上のための定義規定・略称規定における文型定義及びパターンベースの正式名称・略称抽出手法. 言語処理学会 第 31 回年次大会 発表論文集, 2025.
- [9] 北野尚樹, 天笠俊之. 語の出現位置を利用した法令文中の他法令への参照抽出. 第 181 回データベースとデータサイエンス・第 160 回情報基礎とアクセス技術合同研究発表会, 2025.
- [10] 新保彰人, 菅原裕太, 山田寛章, 徳永健伸. 大規模言語モデルを用いた日本語判決書の自動要約. 言語処理学会 第 30 回年次大会 発表論文集, 2024.
- [11] Neel Guha, Julian Nyarko, Daniel Ho, Christopher Ré, Adam Chilton, Alex Chohlas-Wood, Austin Peters, Brandon Waldon, Daniel Rockmore, Diego Zambrano, et al. Legalbench: A collaboratively built benchmark for measuring legal reasoning in large language models. **Advances in Neural Information Processing Systems**, Vol. 36, , 2024.
- [12] John J Nay, David Karamardian, Sarah B Lawsky, Wenting Tao, Meghana Bhat, Raghav Jain, Aaron Travis Lee, Jonathan H Choi, and Jungo Kasai. Large language models as tax attorneys: a case study in legal capabilities emergence. **Philosophical Transactions of the Royal Society A**, Vol. 382, No. 2270, p. 20230159, 2024.
- [13] Jens Frankenreiter and Julian Nyarko. Natural language processing in legal tech. **Legal Tech and the Future of Civil Justice (David Engstrom ed.) Forthcoming**, 2022.
- [14] 金子尚樹. 単語の出現規則に着目した法令中の読み換え規定文の解析手法の提案. 情報処理学会第 85 回全国大会, 2023.