

自己相互情報量と単語群一致度を組み合わせた臨床研究関連法規の類似性評価の検討

大城絢子¹ 岡崎威生² 植田真一郎³

¹ 沖縄国際大学 経済学部 ² 琉球大学 工学部

³ 琉球大学 医学部 臨床薬理学講座

a.ohshiro@okiu.ac.jp okazaki@ie.u-ryukyu.ac.jp

blessyou@med.u-ryukyu.ac.jp

概要

臨床研究は現在3つの指針・省令・法に則っておこなわれている。これらの類似関係を可視化することで、臨床研究の研究者は多忙な診療業務をこなしながらも複数の法令の関係を解釈することができる。本研究では、これらの法令の条文間の類似性を定量的に評価するために、相互情報量と複数の単語(単語群)の一致度・一致単語群の発生頻度・そしてこれらを組み合わせた類似尺度を提案した。類似性推定精度の性能比較実験の結果、既存の尺度に比べ推定精度が改善された組み合わせがあったことを確認した。

1 はじめに

治療法改善や診断ガイドライン検討といった医療の質の向上には、診療現場で生じた臨床的疑問を解決する「臨床研究」が必要である。またその結果は、診療はもちろん患者の予後に影響するため、適切に実施することが重要である。そのため、臨床研究の実施においては既に定められた省令や法令、指針に沿うことが求められる。

例えば1997年には「医薬品の臨床試験の実施の基準に関する省令[1]」(Good Clinical Practice、以後GCP省令とする)が施行され、その内容は多くの研究者に定着していると考えられる。一方で、過去10年においてデータ操作や利益相反に関する問題が発生し、その不適正事案に対する対策として、2015年には「人を対象とする医学系研究に関する倫理指針[2]」(以後、倫理指針とする)の見直し、2018年には「臨床研究法[3]」が施行されたことで臨床研究は法律化され、現在3つの指針・基準・法令のもとにある。そのため臨床研究を実施する研究者はこ

れらの法規に沿った研究を実施することを求められている。しかし、多忙な臨床業務をこなしながら、これらの法規を完全に理解することは容易ではないため、これらの法令を扱う関連部署の支援を受けることも多い。またこのような状況は臨床研究関連の法対応に限ったものではなく、例えば地方自治体で定められる条例など、内容は類似するものの地域により異なる点などは法務の専門家の手作業により比較されている[4]。これらの作業に対し、コンピュータを用いた言語処理のアプローチを導入することにより、作業コストの軽減が期待されている[5][6][7][8]。特に近年では、BERT[9]の文書ベクトル機能やDoc2vec[10]などのDeep Learningを用いた法令間の条文比較などの取り組みも盛んになっている。それにより類似した法律の対応づけなどの取り組みも行われている[11][12]。臨床研究関連の法令においても既述の3つの指針・基準・法令間の関係を可視化することで、診療多忙な臨床研究の研究者が複数の法規の関係を把握することができると期待できる。

本研究では臨床研究関連法規間の関係を可視化するために、各法規の各条文間の全ての組み合わせについての類似性予測の精度を向上させることを目的としている。これまでに、word2vecやtopic modelを用いて関連法規の解釈や法令間の関連性解釈の可能性について考察してきた[13][14]。さらに含まれる共通単語数が多い文章は類似度が高いという仮説のもと、「単語一致度」をもとにこれら3法令の各条文間の類似度を予測した[15]。正誤検証の結果、意味的に「類似性が高い/一致」である組み合わせが正しく獲得された一方で、「類似性が低い/不一致」である組み合わせの条文の類似度が高いとされたケースもあり、課題が残った。

そこで次に、単一の単語ではなく、複数の単語である「単語群」の一致度と複数の単語の同時出現確率を示す自己相互情報量 (Pointwise Mutual Information) を組み合わせて用いることで、各条文間の類似性を予測する尺度を提案した。本稿ではその内容と、既存の類似尺度との性能比較実験の結果について報告する。

2 条文間類似性の定義

本研究の目的は図 1 のように、臨床研究を対象とした既述の法令である L_1 倫理指針、 L_2 GCP 省令、 L_3 臨床研究法の各条文間の類似性を求めることである。

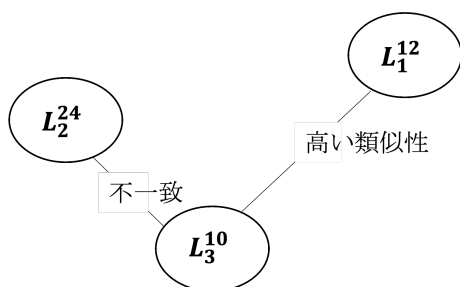


図 1 臨床研究関連法令の条文間の類似性を示した例

図 1 では、例えば L_1 倫理指針の第 12 条と L_3 GCP 省令の第 10 条の類似性が高いことを意味している。本稿では、条文間の類似性について複数の仮説を設定した。複数の単語の組み合わせ (以後、単語群とする) の一致度が高い場合、条文同士の類似性は高くなると考えられる。また一致度に加え、同時に出現する確率 (PMI) の高い単語の組み合わせがある場合、条文同士の類似性は高いと考えることもできる。さらに一致する単語群の発生頻度が高いほど、条文同士の類似性は高いと考えられる。ここで単語群の発生頻度は条文の長さ按比例するため、単語群の発生総数 $F\{x, y\}$ における比率を配慮する必要がある。以上より、これらの仮説を組み合わせ整理すると以下ようになる。

仮説 1 単語群 (2 組) の一致度が高いほど条文間の類似度は高い (Word Group Match : WGM とする)

仮説 2 「単語群の一致度+PMI」が高いほど条文間の類似度は高い (WGM+PMI)

仮説 3 一致した単語群の発生頻度が高いほど条文間の類似度は高い (Frequency Of Word Group Match : FWG とする)

法令 L_a の i 条、法令 L_b の j 条において 2 つの単語 (単語群) $w_{x,y}$ が出現する場合、その相互情報量

$PMI(x, y)$ は式 (1) のように定式化できる。

$$PMI(x, y) = \log \frac{P(x, y)}{P(x)P(y)} \quad (1)$$

また法令 L_a の i 条、法令 L_b の j 条における単語群 $w_{x,y}$ の一致度を $WGM\{w_{L_a}(x, y), w_{L_b}(x, y)\}$ 、とすると、条文間の類似度 $S\{L_a^i, L_b^j\}$ は単語群の発生総数 $F\{x, y\}$ を用いて式 (2) のようにも定式化できる。

$$S\{L_a^i, L_b^j\} = \frac{WGM\{w_{L_a}(x, y), w_{L_b}(x, y)\}}{F\{x, y\}} \quad (2)$$

以上より、本稿では、これらの 3 つの仮説に基づき、条文間の類似度 $S\{L_a^i, L_b^j\}$ について 3 つの類似尺度を定義した。

$$PMI = \frac{PMI(x, y)}{F\{x, y\}} \quad (3)$$

$$PMI+WGM = \frac{PMI(x, y) + WGM\{w_{L_a}(x, y), w_{L_b}(x, y)\}}{F\{x, y\}} \quad (4)$$

$$FWG = \frac{WGM\{w_{L_a}(x, y), w_{L_b}(x, y)\}}{F\{x, y\}} \quad (5)$$

本研究では臨床研究を対象とした法令 3 つ取り上げ、全条文の全ての組み合わせについて類似度を算出した。具体的な手順を以下に示す。

1. 全条文間の全ての組み合わせについて類似度を算出
2. 全類似度の分布をもとに 4 段階の順序尺度「一致/高い類似性/低い類似性/不一致」に変換
3. ベンチマークデータとの順序相関係数を算出し推定精度とする

3 精度比較実験

3.1 用いる法令データ

臨床研究を対象とした法令である倫理指針 L_1 (全 23 条)、臨床研究法 L_2 (全 43 条)、GCP 省令 L_3 (全 59 条) の各法令の全条文の組み合わせ数は、 $\{L_1, L_2\} + \{L_1, L_3\} + \{L_2, L_3\} = 23 \times 43 + 23 \times 59 + 43 \times 59 = 4883$ となる。本研究ではこれらのうち、各法令組み合わせより 30 組選択し、 $30 \times 3 = 90$ の組み合わせについて臨床研究の専門家に評価を依頼しベンチマークデータを獲得した。

3.2 評価指標

提案手法を用いて算出した類似度の精度を、ベンチマークデータを用いて評価した。各条文間の類似

性を評価したベンチマークデータは、臨床研究の専門家にあらかじめ「一致/高い類似性/低い類似性/不一致」の4段階による尺度で評価を依頼し表1のように獲得した。3法令の条文の全組み合わせのうち30組について評価を依頼した。性能比較には、既存の手法として Doc2Vec、単語一致度、Simpson 係数を用いた。

表1 条文間類似度のベンチマークデータ (例)

| L_1 | L_3 | 類似性 |
|-------|--------|-------|
| 第8条 | 第二条 | 不一致 |
| 第8条 | 第十五条の四 | 一致 |
| ⋮ | ⋮ | ⋮ |
| 第2条 | 第二十七条 | 低い類似性 |

類似度の推定精度は、ベンチマークデータとの順序相関係数を用いて [0,1] で評価した。

3.3 比較結果

既存の文章類似度推定手法として単語一致度・Doc2vec・Simpson 係数を挙げ、これらに対して本稿にて提案した類似度尺度である WGM+PMI、PMI、FWG との推定精度を比較した。Simpsons 係数と似た考え方の類似尺度に Jaccard 係数や Dice 係数があるが、本稿では事前実験の結果、精度が高かった Simpson 係数の結果を採用している。3つの法令の組み合わせ{ L_1 倫理指針、 L_2 臨床研究法}、{ L_2 臨床研究法、 L_3 GCP 省令}、{ L_1 倫理指針、 L_3 GCP 省令}の比較結果を表2に示す。

はじめに結果の解釈について述べる。推定精度として用いている順序相関係数が1に近いほどベンチマークデータと正の相関関係にあることを意味し、推定精度が高いと解釈できる。負の値をとりかつ-1に近い場合、類似度がベンチマークデータと負の相関関係になっていることを意味する。これは例えばベンチマークデータにおいて「一致/高い類似性」と判定されている組み合わせに対して「低い類似性/不一致」と推定したことを意味する。

単語一致度を用いた場合、相関係数が負の値や低い正の値をとっていることから、推定精度が低いことを示しており、単語の一致数のみで類似度を求めることが困難であることを示している。Doc2vec も同様のパターンの結果となっている。一致した単語群の発生頻度を類似尺度に適用した FWG を用いた場合、他の手法と比べ推定精度が低く頻度情報のみ用いた場合だと類似性を評価することが困難である

ことを示している。{ L_1 倫理指針・ L_2 臨床研究法}の組み合わせにおいては単語一致群と PMI を組み合わせた「WGM+PMI」や「PMI」を用いた場合、上記の既存の尺度や Simpson 係数の適用時より精度が改善されたことがわかった。また{ L_2 倫理指針・ L_3 GCP 省令}の組み合わせについては他の組み合わせに比べ推定精度が悪いこともわかった。Simpson 係数とのハイブリッド適用などにより、類似尺度の推定精度をさらに改善させる必要があることがわかった。

4 おわりに

自己相互情報量と単語群一致度・単語群発生頻度とそれらを組み合わせた類似尺度を提案し、臨床研究を対象とした3つの指針・省令・法間の条文の類似性を数値化した。性能比較実験の結果、既存の類似尺度より類似度推定精度が改善された組み合わせがあること・既存の尺度との組み合わせによりさらに改善が期待できることを確認した。また性能としてはまだ改善の余地があり、実際の適用レベルを目指すにはまだ改善が必要である。今後の課題として BERT の文書ベクトルや固有表現抽出機能を用いたパターンとの性能比較もおこない、Deep learning や既存の類似尺度とのハイブリッドな手法の構築も試みていく。

表2 各類似尺度による推定精度の比較結果

| 尺度 | 倫理指針・臨床研究法 | 臨床研究法・GCP 省令 | 倫理指針・GCP 省令 |
|------------|---------------------|---------------|--------------|
| 単語一致度 | -0.3705239825 | 0.1030418684 | 0.1823009392 |
| WGM+PMI | <u>0.4535923830</u> | 0.3204137836 | 0.2054946868 |
| PMI | 0.3227211250 | 0.3300901147 | 0.1719666175 |
| FWG | 0.1062887154 | 0.3151868916 | 0.0443173890 |
| Doc2vec | -0.0179077409 | -0.0188153846 | 0.0117952703 |
| Simpson 係数 | 0.3106567258 | 0.3554561002 | 0.2181575929 |

謝辞

本研究は科学研究費補助金 20K18854 の助成を受けたものである。

参考文献

- [1] 医薬品の臨床試験の実施の基準に関する省令(平成9年厚生省令第二十八号)"医薬品の臨床試験の実施の基準に関する省令 | e-Gov 法令検索", (accessed 2020-09-03)
- [2] 文部科学省. 厚生労働省. 人を対象とする医学系研究に関する倫理指針(平成26年12月22日. 平成29年2月28日一部改正) [https://www.mhlw.go.jp/file/06-Seisakujouhou-10600000-Daijinkanboukouseikagakuka/0000153339.pdf (accessed 2020-09-03)]
- [3] 臨床研究法(平成29年法律第16号)"臨床研究法 | e-Gov 法令検索", (accessed 2020-09-03)
- [4] 藤井真知子, "市町村合併における自治体法務の現状と課題: 甲賀市の条例整備を手がかりとして" 龍谷大学大学院法学研究紀要論文, 9号, pp. 181-214, 2007.
- [5] 角野篤泰. スーパーコンピュータを用いた自治体例規の類似度分析と例規データベースへの応用. 名古屋大学法政論集, Vol. 246, pp. 91-69, 2012.
- [6] 竹中要一, 若尾岳志, "地方自治体の例規比較に用いる条文対応表の作成支援" 自然言語処理, Vol. 19 No. 3, pp.194-212, 2012.
- [7] 小山凱丈, 佐野智也, 竹中要一, "明治民法制定時における日仏民法条文の参照関係再推定" 言語処理学会第25回年次大会 発表論文集, pp.398-401, 2019.
- [8] 小山凱丈, 佐野智也, 竹中要一, "明治民法と各国民法との条文類似関係にもとづく立脚点の解析" 言語処理学会第26回年次大会 発表論文集, pp93-96, 2020
- [9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, "Bidirectional Encoder Representations from Transformers",
- [10] Quoc V. Le, Tomas Mikolov, "Distributed Representations of Sentences and Documents", proceedings of the 31st International Conference on Machine Learning, PMLR 32(2):1188-1196, 2014.
- [11] 長裕樹, 中村誠, "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, pp948-951, 2022
- [12] Kaito Koyama, Tomoya Sano, Yoichi, "The legislative study on Meiji civil code by machine learning", Fifteenth International Workshop on Juris-informatics (JURISIN

- 2021)
- [13] 大城絢子, 植田真一郎, "臨床研究を対象とする各法令の特徴抽出とそれらの関連", 電子情報通信学会総合大会, D-5-4, 3月, 2019.
- [14] 大城絢子, 植田真一郎, "テキストマイニングによる臨床研究関連法令の特徴と関連性解釈の可能性", 電子情報通信学会, pp35-pp, 2019
- [15] 大城絢子, 岡崎威生, 植田真一郎, "単語一致度を用いた臨床研究関連法規の関係可視化の検討", 臨床薬理. 2023; 54(1): (in press)