

契約文書における条項マッチング手法の検討

土井 惟成¹ 矢入 健久²¹ 東京大学大学院 工学系研究科 ² 東京大学 先端科学技術研究センター
{n-doi,yairi}@g.ecc.u-tokyo.ac.jp

概要

公共調達においては、発注者ごとに標準の契約条件が異なる傾向がある。そのため、同種の案件の経験が少ない事業者が新規参入を検討する際、契約文書の差異の把握に大きな負担が生じる。本研究では、契約文書間の条項対応を自動推定する条項マッチング手法を提案する。基準となる契約書と発注者の契約書における各条項のテキストに対して類似度行列を計算し、二部グラフのマッチング問題として条項間の対応を推定する。これにより、発注者間の契約文書の差分の把握を容易にすることを旨とする。

1 はじめに

地方公共団体や国立大学法人による公共調達では、発注者が定める標準的な契約文書（約款等）により、当事者間の権利義務、手続、解除条件、支払条件等が規定される。国立大学法人では、このような標準的な契約文書として契約基準を整備している。公共調達に参入する事業者は、これらの条項を読み込んだうえで参入の是非を検討する必要がある。しかしながら、発注者が異なると文書の構成等が部分的に異なる可能性がある。そのため、新規参入者にとって契約文書のレビューの負担は大きいと考えられる。

契約文書の比較の支援には、「複数の契約文書間の条項同士の対応付け」（以下、条項マッチング）が求められる。条項マッチングが安定して行えれば、対応条項間の差分の提示が自動的に実現可能となり、読者の負担が軽減される。

そこで、本研究では、契約文書における条項マッチングを二部グラフのマッチング（割当問題）として捉え、この解決手法を検証する。条項マッチングは、条項間の類似度行列の計算と二部グラフのマッチングで構成される。本研究では、条項間の類似度の計測手法として、TF-IDF[1]、BLEU[2]、ModernBERT[3]系の埋め込みベクトルのコサイン類

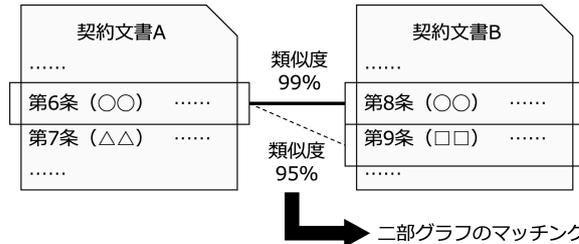


図1 本研究における条項マッチングの概要

似度を扱う。特に ModernBERT 系については複数の日本語モデルを比較し、モデル選択が条項マッチング性能に与える影響も分析する。なお、本研究は試験的な分析として、二部グラフのマッチングにおいては各条項が1対1に対応するものと捉える。条項マッチングの概要を、図1に示す。

本研究では2種類の契約文書を対象とする。試験的な分析として、類似性が特に高いと考えられる地方公共団体の工事請負契約約款を扱う。発展的な分析として、大学間で多様性が大きい可能性がある国立大学法人の役務提供契約基準を扱う。これらの分析を通じて、地方公共団体や国立大学法人による公共調達の契約文書における類似性の傾向について、示唆を得ることを目指す。

本稿の構成は次のとおりである。第2章では関連研究を述べ、第3章では本研究の対象とする契約文書をまとめる。第4章では問題設定を定義し、第5章で提案手法を述べ、第6章で実験について述べた後、第7章で結論と今後の課題を述べる。

2 関連研究

2.1 日本語の法律分野の自然言語処理

法律分野の自然言語処理では、法令や判例等の専門文書を対象に、検索や含意判定といった課題が扱われてきた。日本や他国を対象とした法律文書の情報抽出および含意関係認識を行う国際コンペティションとして、COLIEE が継続的に開催されている

[4]. 法令における用語の検出および校正のように、法律文書の表記等に着眼した応用も報告されている [5]. 日本の法律領域を対象とした事前学習言語モデルの構築や、その性能評価が行われている [6]. また、日本の判決文の論証構造を記述したコーパスの構築 [7] や、法的判断予測のための日本の不法行為事例データセットの構築 [8] が行われている。

これらの研究の対象とする法律分野のテキストは法令や判例が中心であり、日本語の契約文書を対象とした研究は限定的である。

2.2 契約書における自然言語処理

契約書を対象とした自然言語処理では、契約書のレビューの支援を目的とした研究が中心となっている。パブリックドメインの契約書を対象として構築されたデータセットとして、契約条項にラベルを付与したコーパスである LEDGAR が知られている [9]. また、契約書のレビュー支援を目的として、契約書の中で重要となる条項について専門家による注釈が付与されたデータセットとして CUAD が知られている [10]. さらに、契約書の全文を対象とした、自然言語による推論のためのデータセットとして ContractNLI が挙げられる [11].

また、契約書レビュー支援に関する研究としては、利用規約に含まれる不公平な条項を機械学習で検出する手法が提案されており [12], 法的根拠を併せて出力するモデルも検討されている [13]. 複数の契約書を比較する研究としては、複数の言語で提供される利用規約の間における条項の構造差や欠落を検出する手法が提案されている [14].

これらの研究では、主に単一の主体による契約書を対象として、契約書のレビューの支援を目指しているものの、複数の主体による契約書等を比較する研究は限定的である。

3 研究対象の契約文書

3.1 地方公共団体の工事請負契約約款

工事請負契約約款は、個別の工事契約に付随する、工事契約において典型的な条件等を網羅的にまとめた契約文書である。建設業法第 34 条に基づき、国土交通省の中に設置される中央建設業審議会は、標準請負契約約款を作成し、官公庁を含む各当事者へ勧告を行っている。また、中央建設業審議会は、特に公共工事用として、公共工事標準請負契約約款

を作成している。そのため、地方公共団体が用いる工事請負契約約款は、公共工事標準請負契約約款を準用しつつ、各地方公共団体に固有の運用に合わせて、条項の追加・削除や条番号体系の調整が行われていると考えられる。

本研究では、東京都の特別区 (23 区) のうち、工事請負契約約款をインターネット上で公開している 20 自治体を対象とし、各自治体の工事請負契約約款を収集して分析対象とする。特別区は同一の都市圏内にあり、公共工事の発注実務が類似していると推測される。そのため、条項マッチングの難度が低く、試験的な取り組みとして有用だと考えられる。

3.2 国立大学法人の役務提供契約基準

一般的に、国立大学法人は、契約に関する約定事項を定めた契約文書として、契約基準を整備している。文部科学省が所管する工事等の調達に関しては、文部科学省発注工事請負等契約規則において、「工事請負契約基準」「製造請負契約基準」「物品供給契約基準」が定められている。多くの大学は、工事、製造、物品供給、役務提供に関する契約基準がそれぞれ作成されている¹⁾。そのため、工事、製造、物品供給に係る各大学の契約基準は、前述の規則に準じている可能性が高いが、役務提供に係る契約基準は独自に作成されていると考えられる。

役務提供は対象範囲が広く、情報システム、研究支援、施設管理等、業務特性に応じて追加的な条項が必要となり得る。そのため、条項の有無や詳細さの程度、条番号体系が大学間で異なる可能性が高く、地方公共団体の工事請負契約約款よりも条項マッチングの難度が高いと考えられる。

本研究では、多様性が想定される契約文書に対する条項マッチングの難度を検証するため、インターネット上で公開されている 41 件の国立大学法人の役務提供契約基準を分析の対象とする。

本研究の対象とする契約文書の統計情報を、表 1 に示す。

4 問題設定

2 つの契約文書 $A = \{a_1, \dots, a_m\}$ と $B = \{b_1, \dots, b_n\}$ が与えられ、各条項 a_i, b_j は日本語の長文テキストであるとする。条項マッチングでは、すべての条項ペア (a_i, b_j) の類似度 S_{ij} を計算

1) 複数分野の契約基準を包括した文書として契約基準を構成している大学もある。

表1 研究対象の契約文書の統計情報

対象文書	文書数	条項数/文書 (平均)	文字数/条項 (平均)	条項数 (合計)
地方公共団体：工事請負契約約款	20	62.3	430.2	1245
国立大学法人：役務提供契約基準	41	34.9	347.1	1432

して類似度行列 \mathbf{S} を作成し、 \mathbf{S} に基づいて対応集合 $\hat{M} \subseteq \{1, \dots, m\} \times \{1, \dots, n\}$ を推定する。

実際の契約文書では、各条項は必ずしも1対1で対応するとは限らず、1対多または多対1の対応が生じ得る。本研究では、事前の分析により大部分の条項が1対1で対応することが確認できているため、試験的な分析として、各条項は高々1つの条項にのみ対応すると仮定する。そして、 \mathbf{S} を重みとする二部グラフのマッチングとして \hat{M} を求める。参照文書 A は、工事請負契約約款では公共工事標準請負契約約款、役務提供契約基準では東京大学の役務提供契約基準とする。

人手で作成した正解対応集合を G 、予測集合を \hat{M} とすると、対応する条項のペアベースの適合率 P 、再現率 R 、 F_1 は次式のとおり定義する。

$$P = \frac{|\hat{M} \cap G|}{|\hat{M}|}, \quad R = \frac{|\hat{M} \cap G|}{|G|}, \quad F_1 = \frac{2PR}{P+R} \quad (1)$$

5 提案手法

提案手法は、(i) 条項単位のデータセットの作成、(ii) 類似度の計算、(iii) 二部グラフのマッチングで構成される。

条項単位のデータセットの作成においては、契約文書がPDF形式の場合、Adobe AcrobatによりWord形式に変換した後、変換後の文書からテキストを抽出する。その後、抽出したテキストに対してUnicode正規化等を行う。条項分割は、契約文書の特性を踏まえ、正規表現等に基づいて条件を設定する。作成したデータセットは、条番号、見出し、条文で構成される。条より下位の項や号については分割せず、1つの条の中で表現する。

類似度の計算は、見出し（存在する場合）と条文を改行で連結したテキストを入力として、(i) ModernBERT系モデルによる埋め込みベクトルのコサイン類似度、(ii) TF-IDFのコサイン類似度、(iii) BLEU（双方向BLEU平均）を用いる。ModernBERT系埋め込みでは、最終4層の隠れ状態を平均してトークン表現を得る。得られたトークン表現をattention maskを考慮したmean poolingにより1ベクトルに集約し、L2正規化したものを条項ベクトルとして用いる。ModernBERT系モ

デルとしては、cl-nagoya/ruri-v3-130m[15] および sbintuitions/modernbert-ja-130m[16] を比較する。なお、BLEUの値は0-100の範囲で算出されるため、100で割って0-1に正規化した。

類似度行列 \mathbf{S} が得られた後、参照文書 A と比較対象 B の間で1対1対応を満たす対応集合 \hat{M} を推定する。本研究では、 \mathbf{S} を重みとみなした二部グラフ上でマッチングを推定する。この推定手法として、ハンガリー法[17]を採用する。ハンガリー法は、割当問題として最適な1対1対応を厳密に求めるアルゴリズムである。また、実際には「対応する条項が存在しない」条項もあり得るため、マッチングを出力する際には類似度の閾値 τ を導入し、 $S_{ij} < \tau$ の対応は不採用とする。なお、ハンガリー法はコスト最小化問題として解くため、類似度行列 \mathbf{S} をコスト行列 \mathbf{C} に変換し、 $C_{ij} = 1 - S_{ij}$ として線形割当問題を解く。条項数が異なる場合は長方形列として扱う。これは、ダミー条項を追加して正方形列に拡張することと等価である。

6 実験

6.1 実験設定

本章では、提案手法を、(i) 地方公共団体の工事請負契約約款、(ii) 国立大学法人の役務提供契約基準の2種類で評価する。参照文書 A を固定し、比較対象 $B^{(k)}$ との文書対ごとに対応集合 $\hat{M}^{(k)}$ を推定する。工事請負契約約款では公共工事標準請負契約約款を A とし、役務提供契約基準では東京大学の役務提供契約基準を A とする。そのため、役務提供契約基準では、比較対象 $B^{(k)}$ の文書数は40件である。

正解対応集合 $G^{(k)}$ は、2人のアノテータにより人手で作成した。評価は、対応する条項のペアベースの適合率 P 、再現率 R 、 F_1 で行う。集計方法として、文書対ごとの平均（マクロ平均）を計算する。

条項の類似度手法として、ModernBERT系埋め込みのコサイン類似度、TF-IDFのコサイン類似度、BLEUを比較する。ModernBERT系埋め込みは複数モデル（cl-nagoya/ruri-v3-130m および sbintuitions/modernbert-ja-130m）で評価する。

表2 条項マッチングの評価結果 (マクロ平均)

対象文書	類似度手法	適合率 P	再現率 R	F_1
工事請負契約約款	TF-IDF	0.901	0.940	0.920
工事請負契約約款	BLEU	0.981	0.816	0.889
工事請負契約約款	ModernBERT (modernbert-ja-130m)	0.746	0.803	0.774
工事請負契約約款	ModernBERT (ruri-v3-130m)	0.909	0.922	0.915
役務提供契約基準	TF-IDF	0.879	0.918	0.897
役務提供契約基準	BLEU	0.963	0.751	0.841
役務提供契約基準	ModernBERT (modernbert-ja-130m)	0.630	0.670	0.649
役務提供契約基準	ModernBERT (ruri-v3-130m)	0.842	0.857	0.849

TF-IDF および BLEU の算出時における日本語のトークナイズには、形態素解析器として MeCab[18] を使用し、辞書は UniDic[19] とした。BLEU は方向性の影響を避けるため、双方向の平均を類似度として用いる。BLEU の計算には SacreBLEU[20] を使用し、n-gram の次数はデフォルト設定 (4) とした。

マッチング推定としてはハンガリー法を採用し、マッチングを出力する際には類似度の閾値 τ を設け、 $S_{ij} < \tau$ の対応は「対応なし」とする。類似度スコアのスケールは手法ごとに異なるため、 τ は手法ごとに設定する。本研究では、類似度の分布を観察したうえで、ModernBERT 系は $\tau = 0.98$ 、TF-IDF は $\tau = 0.25$ 、BLEU は $\tau = 0.25$ とした。

6.2 結果と考察

条項マッチングの評価の結果を表 2 に示す。

工事請負契約約款においては、TF-IDF が最も高い F_1 (0.920) を示し、ModernBERT (ruri-v3-130m) も僅差 (0.915) で追従した。工事請負契約約款は標準約款の準用が想定されるため、条項の構成や語彙が近く、重要語の一致を捉える TF-IDF でも高い再現率が得られたと考えられる。一方で、文書対ごとには ModernBERT (ruri-v3-130m) が TF-IDF を上回る例も確認されており、条文中の用語選択や言い回しの差、条項の追記位置の差によって表層一致が崩れる箇所では、文脈表現に基づく類似度が誤対応の抑制に寄与する可能性がある。加えて、ModernBERT 系でもモデルにより性能差が大きく、modernbert-ja-130m は他手法より低い F_1 となったことから、条項マッチング用途ではモデル選択が重要であることが示唆される。

一方、役務提供契約基準では、TF-IDF が最も高い F_1 (0.897) を示し、ModernBERT (ruri-v3-130m) (0.849) および BLEU (0.841) が続いた。役務提供は業務内容の幅が広く、大学ごとの規程体系や実務慣行により語彙や構文が変わりやすいことが想定さ

れるが、本実験では見出しや定型語彙の共有が一定程度存在し、語彙一致を利用する TF-IDF の精度が高くなった可能性がある。BLEU は適合率が高い一方で再現率が低くなりやすく、確信度の高い対応候補を提示する用途には有用であるが、網羅的な対応付けには不利となる傾向が示された。ModernBERT 系については、ruri-v3-130m は比較的高い精度を示す一方、modernbert-ja-130m は大きく性能が低下しており、埋め込みモデルの選択が結果に与える影響が大きいことが確認された。

いずれの類似度手法でも確認された傾向として、「発注者の契約解除権」と「受注者の契約解除権」を混同するといった、当事者 (発注者と受注者) の誤りが多く確認された。今後においては、主体と客体の正しい抽出が求められる。

7 おわりに

本研究では、地方公共団体の工事請負契約約款と国立大学法人の役務提供契約基準を対象に、契約文書間で条項の対応を自動推定する手法を検討した。条項間の類似度行列に基づき、ハンガリー法による 1 対 1 マッチングの枠組みにより、両文書種別で高い精度で条項対応を推定できることを確認した。類似度手法の比較から、語彙一致に基づく TF-IDF の精度が高くなる一方、ModernBERT 系埋め込みもモデルによっては近い性能を示すこと、ただし、ModernBERT 系であってもモデル選択により性能が大きく変化し得ることが分かった。

今後は、条項の統合・分割を許す 1 対多対応への拡張や、見出し・参照関係など文書構造情報の活用を進める。また、主体と客体を正確に抽出する手法について検討する。この方法として、各条項を知識グラフで表現することが考えられる。さらに、類似度については複数の手法の併用を検討する。これらに加えて、他の契約文書へ対象を拡張し、より多様な契約文書に対する比較支援の実現を目指す。

参考文献

- [1] Juan Ramos, et al. Using TF-IDF to Determine Word Relevance in Document Queries. In **Proceedings of the first instructional conference on machine learning**, Vol. 242, pp. 29–48. Citeseer, 12 2003.
- [2] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In **Proceedings of the 40th annual meeting of the Association for Computational Linguistics**, pp. 311–318, 07 2002.
- [3] Benjamin Warner, Antoine Chaffin, Benjamin Clavié, Orion Weller, Oskar Hallström, Said Taghadouini, Alexis Gallagher, Raja Biswas, Faisal Ladhak, Tom Aarsen, Griffin Thomas Adams, Jeremy Howard, and Iacopo Poli. Smarter, Better, Faster, Longer: A Modern Bidirectional Encoder for Fast, Memory Efficient, and Long Context Finetuning and Inference. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, **Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 2526–2547, Vienna, Austria, July 2025. Association for Computational Linguistics.
- [4] 狩野芳伸, 吉岡真治, ラベロジュリアーノ, キムミュン, ゴーベルランディ, 佐藤健. COLIEE: 法律文書の情報抽出および含意関係認識を行う国際コンペティション. *人工知能*, Vol. 35, No. 3, pp. 377–384, 2020.
- [5] 山腰貴大, 駒水孝裕, 小川泰弘, 外山勝彦. 事前学習モデル BERT による法令用語の校正. *人工知能学会全国大会論文集*, Vol. JSAI2020, pp. 4P3OS805–4P3OS805, 2020.
- [6] Keisuke Miyazaki, Hiroaki Yamada, and Takenobu Tokunaga. Cross-domain analysis on Japanese legal pretrained language models. In Yulan He, Heng Ji, Sujian Li, Yang Liu, and Chua-Hui Chang, editors, **Findings of the Association for Computational Linguistics: ACL-IJCNLP 2022**, pp. 274–281, Online only, November 2022. Association for Computational Linguistics.
- [7] Hiroaki Yamada, Simone Teufel, and Takenobu Tokunaga. Building a corpus of legal argumentation in Japanese judgement documents: towards structure-based summarisation. **Artificial Intelligence and Law**, Vol. 27, No. 2, pp. 141–170, June 2019.
- [8] Hiroaki Yamada, Takenobu Tokunaga, Ryutaro Ohara, Akira Tokutsu, Keisuke Takeshita, and Mihoko Sumida. Japanese tort-case dataset for rationale-supported legal judgment prediction. **Artificial Intelligence and Law**, Vol. 33, No. 3, pp. 783–807, September 2025.
- [9] Don Tugener, Pius von Däniken, Thomas Peetz, and Mark Cieliebak. LEDGAR: A large-scale multi-label corpus for text classification of legal provisions in contracts. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, **Proceedings of the Twelfth Language Resources and Evaluation Conference**, pp. 1235–1241, Marseille, France, May 2020. European Language Resources Association.
- [10] Dan Hendrycks, Collin Burns, Anya Chen, and Spencer Ball. CUAD: An Expert-Annotated NLP Dataset for Legal Contract Review. **NeurIPS**, 2021.
- [11] Yuta Koreeda and Christopher Manning. ContractNLI: A dataset for document-level natural language inference for contracts. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, **Findings of the Association for Computational Linguistics: EMNLP 2021**, pp. 1907–1919, Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.
- [12] Marco Lippi, Przemysław Pałka, Giuseppe Contissa, Francesca Lagioia, Hans-Wolfgang Micklitz, Giovanni Sartor, and Paolo Torroni. CLAUDETTE: an automated detector of potentially unfair clauses in online terms of service. **Artificial Intelligence and Law**, Vol. 27, No. 2, pp. 117–139, June 2019.
- [13] Federico Ruggeri, Francesca Lagioia, Marco Lippi, and Paolo Torroni. Detecting and explaining unfairness in consumer contracts through memory networks. **Artificial Intelligence and Law**, Vol. 30, No. 1, pp. 59–92, March 2022.
- [14] Andrea Galassi, Francesca Lagioia, Agnieszka Jablonowska, and Marco Lippi. Unfair clause detection in terms of service across multiple languages. **Artificial Intelligence and Law**, Vol. 33, No. 3, pp. 641–689, September 2025.
- [15] Hayato Tsukagoshi and Ryohei Sasano. Ruri: Japanese General Text Embeddings, 2024.
- [16] Hayato Tsukagoshi, Shengzhe Li, Akihiko Fukuchi, and Tomohide Shibata. ModernBERT-Ja. <https://huggingface.co/collections/sbintuitions/modernbert-ja-67b68fe891132877cf67aa0a> (accessed on 2026-01-09), 2025.
- [17] Harold W Kuhn. The Hungarian method for the assignment problem. **Naval research logistics quarterly**, Vol. 2, No. 1-2, pp. 83–97, 1955.
- [18] Taku Kudou. MeCab: Yet Another Part-of-Speech and Morphological Analyzer, 2005. <https://taku910.github.io/mecab/> (accessed on 2026-01-09).
- [19] The UniDic Consortium. Unidic for contemporary written japanese (unidic-cwj). <https://clrd.ninjal.ac.jp/unidic/>, 2021. Version 2021.10 (accessed on 2026-01-09).
- [20] Matt Post. A call for clarity in reporting BLEU scores. In **Proceedings of the Third Conference on Machine Translation: Research Papers**, pp. 186–191, Belgium, Brussels, October 2018. Association for Computational Linguistics.