

# 深層学習モデルは司法試験をどこまで解いているのか： 問題分類とそれに基づく分析

清田直希<sup>1</sup> 藤田真伎<sup>2</sup> 狩野芳伸<sup>1</sup>

<sup>1</sup>静岡大学大学院 総合科学技術研究科 情報学専攻 <sup>2</sup>静岡大学 情報学部  
[nkiyota@kanolab.net](mailto:nkiyota@kanolab.net) [mfujita@kanolab.net](mailto:mfujita@kanolab.net) [kano@inf.shizuoka.ac.jp](mailto:kano@inf.shizuoka.ac.jp)

## 概要

自然言語処理技術の法的応用は重要であるが、テキストの専門性が高く、複雑な構文・意味構造の法律文書を処理するのは依然として難しい。司法試験の自動解答を題材とした COLIEE では、多くの参加チームが end-to-end の深層学習モデルを提出しており、一見高いスコアを獲得しているように見えるが、これらのモデルはブラックボックス的な処理のため、解答の根拠を示すことは難しい。本稿での問題分類とシステム出力の詳細な比較により、現在のシステムは未だ正しい推論過程に基づいた自動解答を行っていないことが示唆された。

## 1 はじめに

近年、自然言語処理技術は急速に発展し、様々な分野での活用が期待されている。司法の分野も例外ではなく、法律文書は文章中の事象を客観的に判断しなければならないことから、自然言語処理の導入に期待が高まっている。

本研究では、毎年開催されている我が国の司法試験の自動解答を行うコンテスト型のワークショップ COLIEE タスクシリーズ[1]-[7]で配布されたデータを対象とする。COLIEE2021[8]で、我々は日本語の述語項に着目した解答器を作成した。その他のチームは、汎用的自然言語処理モデル BERT[9]を使用した機械学習的なアプローチで自動解答を実現していた。

法律文書の処理は、専門性の高い言語が多く用いられているうえ、複雑な構文構造や人物関係、言い換え表現など高度かつ多様な解析技術を必要とする複合的な要素がある。こうした、複雑な文書処理について現状の自然言語処理モデルが論理的な解答を行っているのかを、問題分類を通して分析した。最後に、この分析に基づいて、説明可能な自動解答モ

デルの実現について論じる。

## 2 司法試験の自動解答タスク COLIEE

### 2.1 COLIEE

COLIEE (Competition for Legal Information Extraction and Entailment)は、COLIEE 2014, COLIEE 2015, COLIEE 2016 と人工知能学会国際シンポジア (JSAI-isAI)のひとつである JURISIN (法情報学ワークショップ)において開催してきた。COLIEE 2017 は法と人工知能に関するトップカンファレンスである ICAIL (International Conference of Artificial Intelligence and Law)において開催された。COLIEE 2018 からはカナダ法を用いたタスクが追加されたが、本稿では我が国の司法試験を用いたタスクのみを対象とする。

司法試験は弁護士や裁判官、検察官となろうとする者を対象に行われる試験である。問題形式は択一問題を含む短答式と論文式に分けられて行われる。試験の科目数については、短答式が憲法、民法、刑法の3科目であり、論文式が公法系科目(憲法及び行政法に関する分野)、刑事系科目(刑法及び刑事訴訟法に関する分野)、選択科目(知的財産法、労働法などの中から1科目選択する)の4科目である。

COLIEE ではこのうち民法短答式試験を対象としている。多択を2択にブレイクダウンするとともに、解答に関連する条文を検索する情報検索タスク (Information Retrieval)と、2択を Yes/No で答える含意関係タスク (Entailment)がある。また、知識源として民法の全条文を配布し、問題文共々原文の日本語に加え、英語に翻訳したものを配布している。

COLIEE 2021 では、5つのタスクが用意され、Task1 と 2 はカナダの判例法を用いたタスクであった。Task3 が情報検索タスク、Task4 が含意関係タスクであった。以下で各タスクを定式化する。

情報抽出タスクでは、与えられた問題文  $Q$  を解答するのに関連する民法条文のサブセット  $S_1, S_2, \dots, S_n$  を民法全体から抽出する。

含意関係タスクは、与えられた問題文  $Q$  を同時に与えられた民法条文のサブセット  $S_1, S_2, \dots, S_n$  が含意するかを解答するタスクである。すなわち、含意関係にあれば Yes、そうでなければ No と解答する。

Task5 は司法試験の問題文のみを与えられ、条文検索から解答まで行う Task3 と 4 を組み合わせたタスクとなっている。本稿では Task 4 を対象とする。

なお、Task4 と 5 においては 2 択のブレイクダウンによって、法律文書を正しく理解していなくても、Yes/No で解答をすれば 1/2 の確率で正解になってしまうため、解答の論理性が失われる点が課題とされている。

## 2.2 分析対象の解答器モデル

本稿では我々が COLIEE 2021 において作成した解答器である KIS1[10]を分析対象に用いる。この解答器は我々の COLIEE 2020 のモデル[11]をベースにしており、述語項構造の抽出比較が基盤になっている。本モデルは従来のモデルの節セットに加えて、関連条文中には登場しない表現を抽象表現に置き換え(例: “甲” → “物”), 節分割により節セットの要素を埋められなかった場合、隣接する節の引数を利用して要素を埋める新規節セットを用いている。また、従来では問題文と関連条文の文末の節セットのみが比較対象であったが、本解答器ではすべての節セットを比較することができる項探索モデルを実装した。本解答器は我々が提出したものの中で最も高い精度であった。

また、本稿の分析のためにシンプルな BERT を用いた解答器(Vanilla BERT モデル)を作成した。本モデルは日本語の Wikipedia で訓練されている<sup>i</sup>。事前訓練に使用されたテキストファイルは約 1,700 万文で構成されており、サイズは 2.6GB である。ファインチューニングの際には、COLIEE で出題された過去の問題から作成された 2,164 の訓練データを使用した。

COLIEE 2021 では 80 問の問題を対象に解答の精度を競った。我々が分析対象とするモデルは、述語項ベースの正解が 44 (正答率: .550), vanilla BERT の正解が 50 (正答率: .625)であった。

<sup>i</sup> <https://github.com/cl-tohoku/bert-japanese>

## 3 司法試験の問題分類

司法試験を解答するうえでは、背後にある構文構造、人物の役割と関係性、論理、抽象性など多様で複雑な構造を的確に処理をする必要がある。どの問題解答にどのような要素技術が必要か、平成 29 年度以降のテストデータを対象に解答に必要な要素基準ごとに手作業でマルチラベル分類をした。分類の基準を表 1 に示す。本節ではうちいくつかの分類基準を概説する。

表 1 COLIEE 司法試験問題の技術別分類基準

分類名	分類基準
照応解析	文章の一部を省略している際に、それが何を指しているのかを明確にする際に用いる。「こそあど言葉」のみだけではなく「前文」「同項」なども対象となる
条文検索	条文中に別の条文が指定されている問題
意味役割付与	問題文の中で用いられている名詞が動作主なのか・その動作の対象なのか、手段なのか等を特定する必要がある問題
条件文抽出・判断	問題文・条文のそれぞれの条件文を調べることで正誤が判断できる可能性がある問題
係り受け解析	主語と述語の関係 又は文の並列関係に着目する必要がある問題
箇条書き解析	条文中で箇条書きが用いられている問題。主に箇条書きに書かれているものは条件を満たす変数とみなしてよいと考える。
法律系用語事実関係	民法に記載されている文章と司法試験中の言動についての事実関係を示す必要がある問題
形態素解析	おおむね同じような文章である問題 又は文章を解析しやすくするために特定の格助詞などを入れるのに用いる
準用	準用により、条文の一部を置き換える必要がある問題
否定形解釈	否定形になっていて正誤判定をする問題 又は「～できない(不可能)」「～しなければならぬ(義務)」など「ない」が用いられていてもその意味が異なる問題
一般辞書	法律系の言語ではない一般生活で用いられるような事物が出てきている問題(例: 冷蔵庫, 会社)

一般言換	動詞以外で言い換えがされている問題 (例：一→1)
人物役割付与	複数の人物間の立場や役割を明確にする必要がある問題
人物関係抽出	問題文に出てきた人物(A など)がどの役割(未成年,買主,債務者 等)を担っているのかを明確にする必要がある問題
述語項解析	その問題の動作を明確にする必要がある問題. 特に「～した」「～できる」などはこうした動詞を抜き出すのではなく、その行為を抜き出す必要がある
動詞言換	動詞が言い換えられている問題

### 3.1 照応解析

照応解析は、文章中で登場する「これ」や「この」などの指示内容を明確にする解析である。前述の指示しに加えて、「前章」や「同項」など司法試験特有の指示表現もあり、言語処理においてはその内容を正しく理解する必要がある。

### 3.2 条件文抽出・判断

条件文抽出・判断は問題文及び条文の条件文を調べることで正誤判定ができる可能性がある問題である。例えば以下のような問いである。

<p>関連条文： 第六十条 相続財産に関しては、相続人が確定した時、<b>管理人が選任された時</b>又は破産手続き開始の決定があった時から六箇月を経過するまでの間は、時効は、完成しない。</p> <p>問題文 (正答: No)： 相続財産に関しては、<b>相続財産管理人が選任された場合でも、相続人が確定するまでの間は</b>、時効は完成しない。</p>
--

太字の部分が文章中の条件である。この場合、相続財産の時効不成立の条件が正誤の判断材料になる。関連条文では、管理人選任時から六箇月間は時効が完成しないと記載されているのに対し、問題文では相続人が確定するまでの間と主張されている。このように、これらの文は条件が異なっていることから、答えは No となる。

### 3.3 法律系用語事実関係

法律系用語事実関係とは、法的な知識を要する表現の関係性である。例えば、問題文では「売買」、

関連条文では「双務契約」と記述されていた場合、前者が後者の包含関係にあることを正しく理解しなければ関連条文の内容を適用するかどうかを判断できない。つまり、直接記載されていたり、肯定形や否定形を入れ替えたりしただけの単純な問題ではない。

### 3.4 否定形解釈

否定形解釈は、条文及び問題文の否定形を正しく解釈したうえで、正誤の判断材料にするタスクである。以下のように条文の内容を否定形に反転した司法試験の問題も見受けられる。

<p>関連条文： 第六百七十五条 組合の債権者は、その債権の発生の時に組合員の損失分担の割合を知らなかったときは、各組合員に対して等しい割合でその権利を<b>行使することができる</b>。</p> <p>問題文 (正答: No)： 組合の債権者は、各組合員に対して、その権利を<b>行使することができない</b>。</p>
---

このように、関連条文の内容を否定形にすることで、問題文の主張の正誤も反転する。つまり、この問題の答えは No となる。

否定形解釈は一見、文末の「ない」を認識できれば簡単に評価できるように思えるが、以下のような問いでは「ない」のみで判断することが難しい。

<p>関連条文： 第六百三十三条 報酬は、仕事の目的物の引き渡しと同時に、<b>支払わなければならない</b>。…</p> <p>問題文 (正答: No)： 請負人は、仕事の目的物の引き渡しを要する場合には、これを引き渡した後でなければ、報酬を<b>請求することができない</b>。</p>
---

この場合、条文と問題文それぞれに「ない」が用いられている。しかし、後者は否定形なのに対して、前者は「義務」を表しているため、この問題の答えは No となる。つまり、この問題を解くためには、これらの「ない」に内在される意味を把握しなければならない。

### 3.5 人物関係抽出・人物役割付与

関連条文:

第二百条 占有者がその占有を奪われたときは、占有回収の訴えにより、その物の返還及び損害の賠償を請求することができる。

問題文 (正答: Yes):

A は自己の所有する工作機械を B に賃貸していたが、B は、工作機械の賃貸借契約継続中に工作機械を C に窃取された。この場合、B は、A から独立して、C に対して占有回収の訴えを提起することができる。

先に示した例では、問題文を関連条文と単純に比較することは難しく、正しい解答を出すためには A や B といった人物がどんな関係性であり、それぞれの人物がどのような役割を担っているのかを明らかにする必要がある。この場合、A と B は「賃貸借契約」の関係性である。こうした複数の人物間の関係性を明らかにする必要がある問題には人物関係抽出のタグをつけた。一方で、A は「賃貸人」、B は「賃借人」となり、工作機械を「占有」していることになる。このように、各人物が問題文中でどのような役割を担っているのかを明らかにする必要がある問題が人物役割付与である。

## 4 説明可能な自動解答に向けた分析

### 4.1 問題分類別の正答率

表 2 BERT モデルの問題分類別正答率 (COLIEE2021 抜粋)

技術 #	KIS1[10]	HUKB-2[12]	UA_parser[13]	JNLP.EnssC15 05[14]	OVGU_run3[15]	Vanilla_BERT	COLIEE_baseline	
照応	37	.62	<b>.68</b>	.65	.65	.54	.65	.57
条件	65	.45	<b>.63</b>	.58	.58	.54	.57	.54
法律事実	22	.55	<b>.73</b>	.57	.55	.64	.57	.52
否定	28	.50	.64	<b>.68</b>	.61	.54	.57	.79
人物関係	44	.45	<b>.66</b>	.61	.55	.57	.57	.55
人物役割	34	.44	<b>.71</b>	.65	.53	.53	.65	.59

表 2 は COLIEE 2021 の他参加チームのうち BERT を用いたモデルと、筆者らが作成した Vanilla BERT と、baseline (全て No と解答した場合) とについての問題分類別の正答率を抜粋したものである。BERT モデルは古典的な自然言語処理モデルである筆者らの KIS1 よりも、それぞれの分類で高い正答率であった。

### 4.2 問題分類別の解答状況

本節では、実際に出題されたテストデータを基に、問題の分析を行った (問題例については付録 B を参照)。本問題では先の章で示したように、A や B の役割を明らかにしたうえで、下線部が「債務の目的物の代償である権利の移転」の法律用語系事実関係であることもあり、より難易度が高い問題だと考えられる。使用言語別にみると、日本語モデルは 10/12 に対し、英語モデルは 3/6 であった。しかし本問題では、英語では“a seller (A)”のように役割が明記されており前述のような人物役割の推論過程を省略できる。日本語よりも推論が容易にも関わらず、正答率について大きな差が出た点で現状の推論が不十分なことを示唆できる。

## 5 議論と展望

COLIEE 2021 では 7 割程度の精度を達成した解答器も存在したが、問題分類別の分析を行った結果、同じ技術に分類された問題においても、正答率が安定していなかったため、解答器が法律文書を十分正しく処理するためにはまだ多くの課題があると考えられる。法律分野における実用を想定すると、推論過程とその根拠を示す必要がある。現状の 2 択式で解答器の精度を計る評価方法については、新たな評価尺度を追加し、ブラックボックス的な処理の弊害を除きうる説明可能な技術の開発を進めたい。

### 謝辞

本研究は JSPS 科研費 17H06103 の助成を受けたものです。また、COLIEE のオーガナイザの方々には本研究に使用したデータを提供していただきましたことを厚く御礼申し上げます。

### 参考文献

- [1] “Competition on Legal Information Extraction/Entailment (COLIEE-14), Workshop

- on Juris-informatics (JURISIN) 2014,” 2014. [http://webdocs.cs.ualberta.ca/~miyoung2/jurisin\\_task/index.html](http://webdocs.cs.ualberta.ca/~miyoung2/jurisin_task/index.html).
- [2] M.-Y. Kim, R. Goebel, and K. Satoh, “COLIEE-2015 : Evaluation of Legal Question Answering,” in *Proceedings of the Ninth International Workshop on Juris-informatics (JURISIN 2015)*, 2015, pp. 1–11, [Online].
- [3] M.-Y. Kim, R. Goebel, Y. Kano, and K. Satoh, “COLIEE-2016: Evaluation of the Competition on Legal Information Extraction and Entailment,” in *Proceedings of the Tenth International Workshop on Juris-informatics (JURISIN 2016)*, 2016, pp. 1–13, [Online].
- [4] Y. Kano, M.-Y. Kim, R. Goebel, and K. Satoh, “Overview of COLIEE 2017,” in *Proceedings of the Competition on Legal Information Retrieval and Entailment Work-shop (COLIEE 2017) in association with the 16th International Conference on Artificial Intelligence and Law*, 2017, vol. 47, no. no.Icail, pp. 1–8.
- [5] M. Yoshioka, Y. Kano, N. Kiyota, and K. Satoh, “Overview of Japanese Statute Law Retrieval and Entailment Task at COLIEE-2018,” in *Proceedings of the Twelfth International Workshop on Juris-informatics (JURISIN 2018)*, 2018, pp. 1–12.
- [6] R. Goebel, Y. Kano, M.-Y. Kim, J. Rabelo, K. Satoh, and M. Yoshioka, “COLIEE 2019 Overview,” in *Proceedings of the Competition on Legal Information Retrieval and Entailment Workshop (COLIEE 2019) in association with the 17th International Conference on Artificial Intelligence and Law*, Jun. 2019, pp. 1–9.
- [7] J. Rabelo, M.-Y. Kim, R. Goebel, M. Yoshioka, Y. Kano, and K. Satoh, “COLIEE 2020: Methods for Legal Document Retrieval and Entailment,” in *Proceedings of the Fourteenth International Workshop on Juris-informatics (JURISIN 2020)*, 2020, pp. 1–15.
- [8] R. Goebel, Y. Kano, S. Univesity, M. Kim, and J. Rabelo, “Summary of the Competition on Legal Information Extraction/Entailment (COLIEE) 2021,” in *Proceedings of the COLIEE Workshop in ICAIL.*, 2021, pp. 1–7.
- [9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” Oct. 2018, Accessed: Jan. 08, 2020. [Online]. Available: <http://arxiv.org/abs/1810.04805>.
- [10] M. Fujita, N. Kiyota, and Y. Kano, “Predicate’s Argument Resolver and Entity Abstraction for Legal Question Answering: KIS teams at COLIEE 2021 shared task,” in *Proceedings of COLIEE 2021 Workshop on International Conference on Artificial Intelligence and Law (ICAIL 2021)*, 2021, pp. 15–24.
- [11] R. Hayashi, N. Kiyota, M. Fujita, and Y. Kano, “Legal Bar Exam Solver integrating Legal Logic Language PROLEG and Argument Structure Analysis with Legal Linguistic Dictionary,” in *Proceedings of the Fourteenth International Workshop on Juris-informatics (JURISIN 2020)*, 2020, pp. 148–161.
- [12] M. Yoshioka and Y. Suzuki, “BERT-based ensemble methods with data augmentation for legal textual entailment in COLIEE statute law task,” in *Proceedings of the COLIEE Workshop in ICAIL.*, 2021, pp. 78–83.
- [13] M.-Y. Kim, J. Rabelo, and R. Goebel, “BM25 and Transformer-based Legal Information Extraction and Entailment,” in *Proceedings of the COLIEE Workshop in ICAIL.*, 2021, pp. 25–30.
- [14] H.-T. Nguyen *et al.*, “JNLP Team: Deep Learning Approaches for Legal Processing Tasks in COLIEE 2021,” in *Proceedings of the COLIEE Workshop in ICAIL.*, 2021, pp. 46–53.
- [15] S. Wehnert, S. Viju, D. Shipra, K. Libin, S. Sajjal, and W. de L. Ernesto, “Legal norm retrieval with variations of the bert model combined with TF-IDF vectorization,” 2021.

## 付録 A COLIEE 2021 formal run の問題分類別正答率

Table 1 COLIEE 2021 Task4 formal run の問題分類別正答率 (チーム名はそれぞれ a1: HUKB-1, a2: HUKB-2, a3: HUKB-3, b1: JNLP.Enss5C15050, b2: JNLP.Enss5C15050SilverE2E10, b3: JNLP.EnssBest, c1: KIS1, c2: KIS2, c3: KIS3, d1: OVGU\_run1, d2: OVGU\_run2, d3: OVGU\_run3, e1: TR-Ensemble, e2: TR-MTE, e3: TR\_Electra, f1: UA\_dl, f2: UA\_parser, g: our Vanilla BERT, h: COLIEE\_baseline を表している)

	使用言語	手法	AB問題	照応解析	条文検索	意味役割付与	条件文抽出	係り受け解析	箇条書き解析	法律系事実関係	形態素解析	準用	否定形解釈	一般辞書	一般言い換え	人物関係抽出	人物役割付与	述語項解析	動詞言い換え	TOTAL
#			35	37	9	2	65	22	11	23	21	2	28	21	2	44	34	32	23	80
a1	J	BERT	.63	.68	.67	.00	.60	.73	.55	.65	.81	1.	.64	.57	.50	.61	.68	.75	.61	.69
a2	J	BERT	.66	.68	.67	.00	.63	.73	.55	.70	.81	1.	.64	.57	.50	.66	.71	.81	.70	.71
a3	J	BERT	.63	.65	.67	.00	.60	.73	.55	.61	.81	1.	.64	.52	.50	.64	.65	.78	.65	.69
b1	J	BERT	.54	.65	.44	.50	.58	.50	.55	.52	.76	1.	.61	.57	1.	.55	.53	.69	.61	.64
b2	J	BERT	.60	.57	.56	.00	.58	.64	.45	.52	.71	1.	.64	.57	.50	.57	.59	.69	.61	.64
b3	J	BERT	.57	.59	.44	.00	.58	.59	.55	.52	.76	1.	.57	.62	1.	.55	.56	.66	.61	.64
c1	J	Linguistic	.43	.62	.67	.50	.45	.55	.55	.57	.81	.50	.50	.33	.50	.45	.44	.44	.48	.55
c2	J	Linguistic	.40	.57	.56	.00	.45	.50	.55	.57	.81	.50	.46	.33	.50	.43	.41	.41	.48	.54
c3	J	Linguistic	.43	.59	.56	.00	.45	.50	.55	.57	.81	.50	.50	.33	.50	.45	.44	.44	.48	.55
d1	E	BERT	.46	.41	.44	.50	.43	.41	.73	.35	.48	.50	.54	.33	.50	.41	.44	.50	.35	.44
d2	E	BERT	.46	.57	.56	.50	.51	.55	.64	.48	.67	1.	.43	.48	.50	.52	.47	.50	.57	.55
d3	E	BERT	.51	.54	.67	.50	.52	.59	.64	.57	.62	1.	.54	.38	.00	.57	.53	.59	.70	.59
e1	J	T5	.54	.51	.44	.50	.51	.45	.55	.48	.71	1.	.68	.57	1.	.59	.56	.47	.70	.59
e2	J	Multee	.54	.51	.56	.00	.51	.59	.45	.43	.71	1.	.57	.48	1.	.55	.56	.59	.74	.59
e3	J	Electra	.51	.49	.44	.50	.48	.32	.55	.43	.57	.50	.57	.62	.00	.52	.53	.34	.48	.50
f1	E	BERT	.46	.46	.33	.50	.45	.36	.55	.43	.76	.50	.50	.48	1.	.50	.50	.47	.57	.54
f2	E	BERT	.60	.65	.44	.00	.58	.59	.55	.57	.81	1.	.68	.57	.50	.61	.65	.69	.83	.68
g	J	BERT	.66	.65	.56	1.	.57	.50	.73	.57	.76	1.	.57	.67	.50	.57	.65	.66	.70	.63
h	-	No to All	.60	.57	.67	.50	.54	.55	.64	.52	.24	1.	.79	.52	.50	.55	.59	.69	.74	.53

## 付録 B

問題文 (正答: Yes):

A と B は, A が所有する骨董品甲を B に 100 万円で売却する旨の売買契約を締結した. 売買契約の締結後, B が代金 100 万円を支払ったが, A が甲を B に引き渡す前に, 甲が B の責めに帰すべき事由により焼失した場合において, A が甲の焼失による損害をてん補するために支払われる損害保険金 70 万円を得たときは, B は, A に対し, 70 万円の支払を請求することができる.

(If, after the conclusion of the contract for sale, a buyer (B) pays 1000000 yen for X, but before a seller (A) delivers X to B, X is burned-out due to grounds attributable to B, and A obtains 700,000 yen of the insurance money for damage caused by the destruction of X, B may claim payment of 700,000 yen to A.)

関連条文:

第四百二十二条の二 債務者が, その債務の履行が不能となったのと同じ原因により債務の目的物の代償である権利又は利益を取得したときは, 債権者は, その受けた損害の額の限度において, 債務者に対し, その権利の移転又はその利益の償還を請求することができる.

(Article 422-2 If the obligor acquires a right or profit as a substitute for the subject matter of the obligation due to the same cause as the one that has rendered the performance of the obligation impossible, the obligee may demand the transfer of the right or reimbursement of the profit from the obligor, to the extent of the amount of damage sustained thereby.)