

司法試験自動解答を題材にした BERT による法律分野の含意関係認識

星野 玲那¹ 狩野 芳伸²

静岡大学 総合科学技術研究科

rhoshino@kanolab.net¹, kano@inf.shizuoka.ac.jp²

1 はじめに

現在、法律分野での自然言語処理技術の応用が注目を集めている。司法の判断は専門家でないとなし、また過去の大量の判例から関連事件を探したり、供述書から重要な事実を探したりするのに多大な労力がかかっている。専門知識を既存のテキストデータから自動的に体系化する技術があれば、そのような作業を補助することができる。将来的には過去の判例から自動的に前提知識を学習し、裁判で人の判断を補助するシステムの実現が理想である。

本研究の目的は、法律文書の表す内容を機械的に構造化して自動的に学習し、民法条文などと比較して法と適合しているか意味的に判断することを可能にするシステムの開発である。そうしたシステムを開発するためのベンチマーキングとして、我が国の司法試験を題材に自動解答を競う COLIEE (Competition on Legal Information Extraction/Entailment) シリーズがある。我々自身の先行研究として、ICAAIL 2019

(International Conference on Artificial Intelligence and Law 2019) で行われた COLIEE 2019 (Competition on Legal Information Extraction/Entailment) の Statute Law competition[1]にて、実際に出题された司法試験の民法に関する問題を内容が正しいかどうか Yes/No の 2 択で答える短答式の問題文を解答し、正答率を競うタスクに参加したものがあつた。その際にはルールベースのシステムを使用して正答率は 2 位の結果であつた。本稿では、新しく BERT モデル[2]を使用した解答器を開発し、専門知識が求められる法律文書で BERT を使用する際の手法について考察する。

2 タスク説明

COLIEE は 2014 年から JURISIN (Juris-Informatics) のワークショップの一つとして開催されており、タスクの名称や内容は年度によって異なる。本稿では COLIEE 2019 の Task 4 を対象とし、以下では単に COLIEE 2019 とよぶ。

問題文:

15歳に達した未成年者は、遺言の証人となることができる。

関連条文：第九百七十四条

次に掲げる者は、遺言の証人又は立会人となることできない。

一 未成年者

二 推定相続人及び受遺者並びにこれらの配偶者及び直系血族

三 公証人の配偶者、四親等内の親族、書記及び使用人

図 1 Task4 の問題例 (正解 : Yes)

Task4 では司法試験の民法短答式問題から抽出された問題文と、その問題を解くのに必要な知識が書かれている条文が与えられ、その問題文が正しいかどうか Yes/No で解答する。図 1 が実際の問題例で、t1 が問題を解答するのに必要な条文であり、t2 が問題文である。日本語の民法条文・問題文を翻訳した英語版テキストも配布されたが、本研究では日本語版のみを用いた。

問題は法律の専門家が選定しており、関連条文だけで解ける問題のみを選ぶため、外部知識は必要ないとしている。しかし、だれでも外部知識なしに解けるわけではなく、法律について学んだことのない人には解けない問題もある。法律について学んだことのない日本人大学生が平成 26 年度の問題を解いたところ、8 割程度の問題に正解した。2 割程度の問題は一般的な日本語が理解できるだけでは解けないといえる。その理由としては、専門用語が存在すること、関連条文が複数個ある場合に適用すべき部分を見つけにくいこと、並列関係を表すときなどに法律文書独特な言い回しが使われるため、その言い回しについて理解しておかなければ意味を理解しにくいことが考えられる。

3 先行研究

COLIEE 2019 で正答率 1 位だった Kim ら[3]は、日本語の問題を機械翻訳で韓国語にすることで文章構造を平易にして解析しやすくするという手法をとっている。システムの流れとしては、問題文と関連条文の文章の構造を分析した後、「もし」「～したとき」などの単語が入っていた時は条件部分、文末部分を結論部分として文章を分割し、関連条文が問題文の条件部分に関するものか結論部分に関するものかを分析している。さらに論理判定に重要な否定表現を接頭辞だけでなく単語事態が否定形の意味を表しているかどうかを辞書によって検出している。最終的には文章構造や否定表現の分析結果を使用してルールベースのシステムと Support vector machine (SVM) を使用した機械学習のシステムで解答を決めている。

また、Task4 で深層学習を使った例としては、Baban ら[4]が英語版の問題において BERT モデルを使用した解答器を提案している。Baban らは問題文と関連条文をそれぞれ二つの文とみなして、文分類タスクとファインチューニングを行っている。その中で、BERT モデルの 12 層のレイヤーを通した後、全結合層に入力するだけでは精度がでなかったため、代わりに BERT モデルの出力ベクトルを機械学習モデルである SVM や eXtreme

Gradient Boosting (XGBoost) [5]に入力して分類するという手法をとっている。

我々はこれまでの COLIEE に 2016 年から 2019 年度までの 4 回に参加している。COLIEE 2016[6]では格文法の理論を利用して述語項解析を用いた解答システムを開発し[7]、COLIEE 2017[8]で文章の比較について改良を行った。[9]COLIEE 2018[10]では述語項解析において課題となった類義語に対応するために類義語辞書を作成し、また文字列の表層的表現の曖昧性をなくすことを目的として、述語項解析に FrameNet を導入することによって正答率が向上した[11][12]。COLIEE 2019 では条文の例外に着目して論理を判定する機能を追加した[13]。本稿では、2019 年度のシステムの結果を利用した新しい手法を考案する。以下に、2019 年度のシステムの概要を示す。

3.1 述語項解析

述語項解析によって文中の述語に注目して文を構造化し、内容を簡潔に表現することで複数の文を単純に比較可能にすることを目的とする。具体的には、単語間の関係から文を一つの動作を表す意味的なまとまりで区切ることで構造化する。単語間の関係の表現には表層格を用いる。そして、一つの述語とその述語にかかる文節の集合を節と定義し、節を展開節、条件節とその他の三種類に分類した。展開節は文末の述語を含む節である。文の最後に出現する述語がその文の内容の全体に影響する、文の主題を表現しているという考え方によるものである。そのような展開節の内容に対して条件を表現する節が条件節である。「～ときは」「～場合は」のような特定のフレーズを含む節を条件節としている。

文章を比較するときは、節の中心の述語、その主語と目的語の 3 つのセットを用いて、問題文と条文の展開節同士、条件節同士を比較する。基本的には、一文が複数の条件節と 1 つの展開節に分ける。一つの条文は複数文であるため、問題文と比較するときは一文ずつ比較していく。

3.2 システム全体の構造

システム全体は違う機能を持ったモジュールの集合となっている。問題文と文章を比較する際にどの程度単語が一致していたら同じ内容として扱うかなどの条件によって違う機能を持ったモジュールを作り、モジュールごとに解答を出す。そして、その解答を集計することで最終的な解答を出す仕組みとなっている。結果の集計ではモジュールに優先順位をつけ、優先順位が高いモジュールの答えを最終結果とする方法と、問題文と条文を比較するときに文章がどれほど一致しているかを表す確信度を素性として、SVM による学習と推定を行った。

4 提案手法

問題を解くのに必要なスキルを人手で分類した結果、条件文が重要な問題が出題されることが多いが、前述の我々のシステムはその種の問題の正答率が低かった。また、条文外の専門知識がなければ解けない問題が存在するため、文同士の単純比較だけでは対処できないことが課題であった。

こうした課題に対応するために、BERT をベースとする深層学習を用いたシステムで、文同士の関係性や単語の意味を学習し、解答までを行うモデルを提案する。

4.1 事前学習

現在、BERT の日本語事前学習済みモデルが複数公開されている。¹²しかし、これらの日本語訓練済みモデルを司法試験の問題に適用するには複数の問題点があった。まず、事前学習の入力系列長が事前学習に比べて短すぎることがある。今回のタスクの問題文は基本的に一文だが、複数の人物や対象物を説明するために一般的な文より文長が長い。また、問題を解くのに関連条文が必要となるが、問題文の後ろに関連条文を繋げる方法では入力文が 3 文以上になることが多い。さらに法律文書で使われる専門用語が出現した場合、形態素解析や一般的な表現が使われた文書で学習したトークン分割手法では一単語が複数のトークンに分割されがちで、トークン数が多くなる傾向がある。これらの要因によって、転移学習を行う際の入力系列長が事前学習より長くなることが多いが、系列長の違いが大きすぎると 2 つの問題が起きる。

第一に、BERT では入力系列長の限界付近・文の後ろの方になるほど学習が進まず、転移学習で長い文を入力とした時、事前学習をうまく生かせない可能性がある。1 つの入力文の系列が入力系列長の限界に達しないとき、入力文の終端から入力系列長の限界までパディングを行うことで全ての入力系列の長さを揃える。そして学習時にはパディングされた部分をマスクして学習が行われないようにする。このとき、事前学習で短い文ばかりで学習を行っていた場合、系列の後ろの方は学習で利用されないため、長い文を入力した場合の学習が不十分となる可能性がある。

第二に、事前学習済みモデルの入力系列長の限界が事前学習データセットの長さに合わせているため、転移学習で長い文を入力すると文の途中で打ち切らざるをえない。一つの問題に対して解くのに必要な条文は複数ある可能性があり、BERT の複数文の文脈を考慮することが可能な長所を生かすために関連条文全文を入力すると入力長はさらに長くなる。日本語では文末表現が文全体の意味を変えうるため、入力系列限界で文が途切れるのは好ましくない。前述の訓練済みモデルでは形態素単位で限界が 256 であるなどの例があったが、今回のタスクでは不足すると判断した。

¹ <http://nlp.ist.i.kyoto-u.ac.jp/index.php?BERT> 日本語 Pretrained モデル

² <https://github.com/yoheikikuta/bert-japanese>

モデル名 (条件)	wikipedia (L: 6)	legal (L: 8)	legal (L: 8, 複数文禁止)	wikipedia+legal (L: 8)	wikipedia+legal (L: 6)	wikipedia+legal (L: 6, 語彙数: 32000)
全結合層	0.58	0.6	0.55	0.62	0.58	0.56
XGBoost	0.53	0.58	0.53	0.5	0.5	0.56
XGBoost+module	0.54	0.61	0.54	0.48	0.47	0.54

表 1 Task4 の結果(事前学習の方法ごとの正答率) ※L はレイヤー数、複数文禁止は次文予測にて後ろの文を 1 文にする、語彙数は注記がなければ 16000、COLIEE 2019 での我々のルールベース解答器 (KIS_module) の結果は 0.62

そこで我々は、オリジナルの BERT の設定に変更を加えた。まず、事前学習で行う次文予測で予測する 2 文目について、複数文がありえるようにした。これは前述した系列長の違いによって起こる問題に対処するためである。長い入力に事前学習時点から条件を合わせることで、問題文と関連条文を合わせた入力にも対応できる言語モデルが学習可能になると考えられる。また、次文予測の際に次文ではない負例を作るとき、BERT では他の文書から抽出した文を使うことがあるが、BERT の後に発表された ALBERT[14]を参考に、同じ文書内からのみとした。

4.2 転移学習

転移学習を行う際に課題となるのが、学習データの少なさである。平成 18 年度から平成 28 年度までの問題で利用可能なもの全 578 問を訓練で使用するが、深層学習を行うには少量である。BERT では、数百メガバイト以下のデータでは損失が下がらず、学習が収束しない傾向にある。対策として、隠れ層の数やレイヤーの数を減らし、さらに Baban ら[4]を参考に BERT からの出力を XGBoost の入力として使用し予測を行うことにした。全結合層を使わないことで、少ないデータでも学習できるようにすることを目的としている。

また、XGBoost に BERT の出力に加えて、前述した COLIEE 2019 のルールベースの解答器で問題を解くときに生成した中間的な情報を渡すことも試みた。解答器は問題文から各種処理を行うので、問題文を含まない事前学習の文書には適用できないが、事前学習を必要としない XGBoost の入力には BERT の出力と同時にルールベースの解答器の中間情報を数値として渡すことができる。そこで、XGBoost の入力にモジュールごとの解答と確信度を追加した。モジュールごとの解答と確信度には格解析や述語項解析の結果などを複数の方法で織り交ぜた情報が含まれており、BERT のみでは学習の難しいモジュールの情報を学習することが可能ではないかと考えたからである。

5 実験

COLIEE データの正答率でシステム全体の評価を行う。開発には平成 18 年度から 28 年度まで 10 年分の問

題を使用し、最終評価に平成 29 年度の問題を使用した。年度ごとに問題数は異なっており、平成 18 年度から 28 年度までが合計 717 問、29 年度が 98 問であった。

述語項解析での文章の解析には日本語形態素解析器 JUMAN Ver. 7.01³と日本語構文解析器 KNP Ver. 4.19⁴を使用し、JUMAN に日本語法律用語辞典「有斐閣法律用語辞典第 4 版」[15]をユーザ辞書として 12,612 単語を追加した。

言語モデルを一般的な日本語の文書で学習した場合と法律文書を使用した場合を比較するために、BERT の事前学習には日本語 Wikipedia と裁判所の判例検索ページで取得できる最高裁判所の判例を用いた。判例検索では高等裁判所や下級裁判所の判例も公開されているが、最高裁判所のみ民事と刑事を選択して検索可能なため、今回は最高裁判所の民事事件の判例だけを使用した。取得した判例は 12575 件で、PDF からテキストのみを抽出して使用した。ルールベースの解答器への入力と BERT の転移学習には司法試験の問題と民法条文全 1044 条を使用した。Wikipedia は全ページの 10 分の 1、117,510 ページを利用した。

BERT を訓練する際に文章をトークン化するには Sentencepiece[16]を用いた。訓練データを用いて Sentencepiece モデルを語彙数 16000 で学習し、サブワードに分割した。Wikipedia を BERT の事前学習の訓練データとするときは Wikipedia のみで、判例を用いるときは判例のみで学習した。

BERT の訓練時のハイパパラメータは特記がない限りバッチサイズ: 26, 語彙数:16000, 隠れ層:256, オプティマイザ: Adam, 学習率: 1e-3, 最大系列長: 512, ヘッド数(Multi-head Attention): 8, Dropout 率: 0.1 とし、レイヤー数(L)を変えて結果を比較する。

XGBoost を学習する際は、最大ラウンド 100 まで学習を行い、訓練データでの正答率が最も高かったモデルを使用する。

6 結果と考察

表 1 が Task 4 の結果である。モデル名の wikipedia は事前学習で Wikipedia を使ったこと、legal は最高裁の判例、法律文書を使ったことを示す。また、

³ <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

⁴ <http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP>

	問題数	wikipedia	legal (XGBoost)	legal (XGBoost+module)	wikipedia+legal (L: 8)	KIS_module (参考)
条件節判断	83	49 (0.59)	45 (0.54)	50 (0.6)	51 (0.61)	52 (0.62)
人物役割	66	36 (0.54)	36 (0.54)	36 (0.54)	38 (0.57)	43 (0.65)
人物関係	66	36 (0.54)	36 (0.54)	36 (0.54)	35 (0.53)	43 (0.65)
法律用語	24	16 (0.66)	11 (0.45)	16 (0.66)	16 (0.66)	13 (0.54)
動詞言換え	21	15 (0.71)	16 (0.76)	17 (0.80)	14 (0.66)	10 (0.47)

表 2 解答に必要な能力ごとの正答数 (正答率)

XGBoost+module は XGBoost の入力にルールベースのモジュールの情報を入力している。最も正答率が高かったのは次文予測を複数文にして、事前学習に最高裁の判例と Wikipedia を用いて、最終層を全結合層にしたモデルで、COLIEE2019 で我々のルールベース解答器と同じ正答率であった。

事前学習で使用する文書については、Wikipedia だけを使うよりも法律文書を加えたほうが精度が上がる傾向がみられた。BERT の一般的なねらいは複数タスクで応用可能な汎用言語モデルを学習することだが、文書の種類、前提とする知識が違う場合はタスクに特化した事前学習データが有効なことがあると考えられる。

次文予測で複数文を許すか許さないかの比較では、許したほうが正答率が高かった。事前学習から入力の文長を長くしたことで、長文になりやすい関連条文に対応可能な言語モデルを学習したと考えられる。

人手で問題の解答に必要な工程を割り出し、解答するのにどの要素技術が必要かを分類したうえで、分類ごとに正当性を計算した結果を表 2 に示す。

KIS_module は COLIEE 2019 の我々のチームで最も成績が良かったルールベースのシステムの結果である。

今回のモデルは問題数が多い条件節判断や人物関係に関する問題の正答率が低く、これが全体の正答率が伸び悩んだ理由であると言える。細かい条件や論理判断を行うのはルールベースは人の手による調整が利くため得意なのではないかと考えられる。その反面、法律用語の知識や動詞言換えが必要な問題の正答率が高い。これは、表面的な文面の差異を言語モデルの学習によって吸収することができた結果ではないかと考える。

7 おわりに

我々は COLIEE の司法試験自動解答タスクにおいて、新しく BERT モデルを使用した解答器を開発し、専門知識が求められる法律文書で BERT を使用する際の手法について提案した。BERT の事前学習で法律文書を用いて言語モデルを学習することで、法律用語の知識や動詞言換えなど文面の表層的な表現について対応できるモデルが可能となった。

司法の判断において重要な、文章の条件部分や人物関係の把握についてはルールベースのシステムのほう

が正答率が高かったため、改善の余地があると考えられる。

参考文献

- [1] R. Juliano, M.-Y. Kim, G. Randy, M. Yoshioka, K. Yoshinobu, and S. Ken, "COLIEE 2019 Overview," *Proc. Sixth Int. Compet. Leg. Inf. Extr. (COLIEE 2019)*, pp. 1–9, 2019.
- [2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2018.
- [3] M.-Y. Kim, J. Rabelo, and R. Goebel, "Statute law information retrieval and entailment," *Proc. 17th Int. Conf. Artif. Intell. Law, ICAIL 2019*, pp. 283–289, 2019.
- [4] G. Baban, B. Dibyanayan, S. Tanik, and E. Asif, "IITP in COLIEE@ICAIL 2019: Legal Information Retrieval using BM25 and BERT," *Proc. Sixth Int. Compet. Leg. Inf. Extr. (COLIEE 2019)*, pp. 45–49, 2019.
- [5] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 13-17-August-2016, pp. 785–794, 2016.
- [6] M.-Y. Kim, R. Goebel, Y. Kano, and K. Satoh, "COLIEE-2016: Evaluation of the Competition on Legal Information Extraction and Entailment 2 - The Legal Question Answering Task," 2016.
- [7] R. Taniguchi and Y. Kano, "Legal yes/no question answering system using case-role analysis," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10247 LNAI, pp. 284–298, 2017.
- [8] Y. Kano, M.-Y. Kim, R. Goebel, and K. Satoh, "Overview of COLIEE 2017," vol. 47, no. Icail, pp. 1–8, 2017.
- [9] Y. Kano, R. Hoshino, and R. Taniguchi, "Analyzable Legal Yes/No Question Answering System using Linguistic Structures," *4th Compet. Leg. Inf. Extr. Entailment (COLIEE 2017), 16th Int. Conf. Artif. Intell. Law (ICAIL 2017)*, vol. 47, pp. 57–67, 2017.
- [10] M. Yoshioka, Y. Kano, N. Kiyota, and K. Satoh, "Overview of Japanese Statute Law Retrieval and Entailment Task at COLIEE-2018," *Twelfth Int. Work. Juris-informatics (JURISIN 2018)*, pp. 1–12, 2018.
- [11] R. Hoshino, R. Taniguchi, K. Naoki, and Y. Kano, "Question Answering System for Legal Bar Examination Using Predicate Argument Structure," pp. 207–220, 2019.
- [12] R. Taniguchi, R. Hoshino, and Y. Kano, "Legal Question Answering System Using FrameNet," pp. 193–206, 2019.
- [13] R. Hoshino, N. Kiyota, and Y. Kano, "Question Answering System for Legal Bar Examination using Predicate Argument Structures focusing on Exceptions," *Proc. Sixth Int. Compet. Leg. Inf. Extr. (COLIEE 2019)*, pp. 38–44, 2019.
- [14] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soicrut, "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations," 2019.
- [15] 法令用語研究会, 有斐閣 法律用語辞典, 第4版. 有斐閣, 2012.
- [16] T. Kudo and J. Richardson, "SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing," *EMNLP 2018 - Conf. Empir. Methods Nat. Lang. Process. Syst. Demonstr. Proc.*, pp. 66–71, 2018.