

司法試験自動解答におけるルールベースと機械学習のシステム構築とアンサンブル

藤田真伎¹ 狩野芳伸¹

¹静岡大学 情報学部

mfujita@kanolab.net

kano@inf.shizuoka.ac.jp

概要

法律分野における自然言語処理技術の利用は、大きく期待されている応用分野の一つである。究極的な目標は裁判過程の自動化であり、裁判のコストの高さや、判決を市民に説明するといった課題に対応できる可能性がある。法律分野の自然言語処理技術に関するワークショップとして、我が国司法試験の民法分野短答式問題の自動解答を行い性能を競う COLIEE が毎年開催されてきた。我々は COLIEE の含意関係認識タスクを対象に、ルールベースのシステムと機械学習ベースのシステムの実装、改良を行い、さらにこの二つのシステムのアンサンブルを行い、解答性能の向上を確認した。

1 はじめに

法律分野の自然言語処理技術に関するワークショップとして、我が国司法試験の民法分野短答式問題の自動解答を行い性能を競う Competition on Legal Information Extraction and Entailment (以下 COLIEE) [1]が毎年開催されてきた。COLIEE の Task 4 は民法条文一覧と過去問題が与えられ、それをもとにシステムを作成し、テスト問題の解答の精度を競うものである。このタスクは、図 1 のように問題文とその問題を解くのに必要な関連民法条文の二つが与えられ、問題文が正しいか否かを Yes か No の二択で答え、解答の精度を競うものである。

COLIEE 2021 の Task 4 に、我々はルールベースの解答方法を用いるシステム [2]を使用した。他のチームは BERT [3]を中心に使用するチーム [4] [5] [6]や T5 をベースにアンサンブルを行うチーム [7]、グラフニューラルネットワークを用いるチーム [8]など機械学習ベースのシステムを使用しており、ルールベースのシステムを使用するチームは我々のみだった。推論過程がブラックボックスであ

る機械学習ベースのシステムと異なり、ルールベースのシステムは人間の思考に沿った推論を行うことができたため、究極的な目標である裁判過程の自動化の際に重要となる、説明可能な出力といった面で優れている。

本論文では、ルールベースのシステムの改良を説明し、また BERT を用いた機械学習ベースのシステムの実装、改良を行う。その後、二つのシステムのアンサンブルを行った結果、精度が向上した。このことから、ルールベースのシステムは説明可能な出力が可能だけでなく、精度向上にも貢献できることが確認できた。

<正解ラベル:No>

<関連民法条文>第九十二条 取引行為によって、平穩に、かつ、公然と動産の占有を始めた者は、善意であり、かつ、過失がないときは、即時にその動産について行使する権利を取得する。

<問題文>強迫を受けてした動産売買契約を取り消した売主は、取消し前に買主から当該動産を善意かつ無過失で買い受けた者に対して、所有権に基づいて、当該動産の返還を求めることができる。

図 1 COLIEE Task 4 (含意関係認識) の問題例

2 関連研究

ルールベースのシステムの先行研究として星野 [9]のものがあり、これをベースに用いる。これは文章を構文解析システム [10]で解析し、解析結果から節分割をする。その後、解析結果を再度使用し、各節で主語、述語、目的語の節セットを抽出して解答を行うものである。このシステムは複数のモジュールに分かれており、モジュールにより解答方法と精度が異なる。詳細探索モジュールは関連民法条文と問題文節セットが全て一致していたら比較を行う、曖昧探索モジュールは節セットのうち二つが

一致していたら比較を行うなど比較条件が異なっている。

2021年に開催されたCOLIEE 2021において、Task 4ではBERTを用いた北海道大学のチーム [4]が最高の成績を収めた。このチームは、主に訓練データの拡張とBERTのアンサンブルによって性能向上を図っている。訓練データの拡張には、配布される民法条文一覧を使用している。文章を分割することでデータ数を増やし、またその文章の一部を否定形にすることで論理を反転させることでデータ拡張をしている。また複数のBERTのアンサンブルも行っており、異なるデータで学習させたモデルの中で、精度の高いものを複数選択し結果をアンサンブルすることにより精度を向上させた。

3 提案手法

3.1 ルールベースのシステムの改良

3.1.1 モジュールの構成

我々が以前に構築したルールベースのシステム [9]を改良し、精度の向上を図る。モジュールの詳細は次項以降で説明する。モジュールごとに解答できる問題と正答率が異なるため、正答率の高いモジュールから順に解答を行い、最後に全ての問題に解答できるが正答率の低い全探索モジュールで解答を行っている。全探索モジュールで多くの問題に解答してしまい、全体の正答率が下がるのが問題点となっていたため、多くの問題に解答でき正答率の高い中間層の項探索モジュールを作成し、精度向上を図った。図2がモジュールの構成図である。

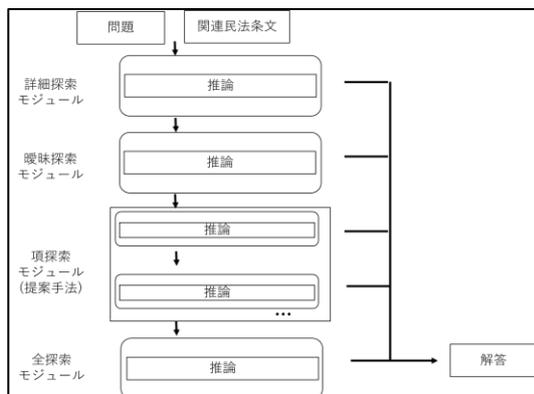


図2 ルールベース改良後モジュール構成図

3.1.2 二重構文解析

既存のシステムでは、文章を一度構文解析し、文章を複数の節に分割した後、その構文解析結果を再度使用し、節ごとの主語、述語、目的語の節セットを作成していた。しかし、この方法では文章が長文の場合に節セットが上手く抽出されない場合や、入れ子構造に対応できないことがあった。そのため、節分割後に再度構文解析を行い、その結果を節セット抽出に使用することで上記の問題に対応した。この手法は全てのモジュールで使用する。

3.1.3 新規節セット

先述の二重構文解析を用いても、節セットが全て抽出されない場合があった。そのような場合に備え節セットの抽出方法を改善した。構文解析結果を使用しても節セットが全て抽出されなかった場合、助詞の「は」や「が」の前に位置する単語を主語に、助詞の「に」と「を」の前に位置する単語を目的語に設定した。また、比較しやすい節セットの作成のため、人物名を全て「人」に、「甲」「乙」「丙」を全て「物」に置き替える処理を追加し、比較を行いやすい節セットを作成した。

3.1.4 項探索モジュール

先述の新規節セットを用いた新たなモジュールを作成する。このモジュールを項探索モジュールとした。このモジュールでは新規節セットと既存の節セットの両方を使い、節セット内のどれか一つでも一致したら比較対象とみなし解答を行う。複数の条文が一致した場合、一致数が多いものを優先し、数が同じ場合は精度が高いと考えられる既存の節セットで一致したものを優先する。

3.2 機械学習ベースのシステムの実装

機械学習手法としてBERTを用いた。14年分の過去問題のうち2年分を検証データ、残りを訓練データとして使用したものを7回実行し、全ての過去問題を検証に使用した。入力の最大文字列長は512とし、それ以上の長さの問題は解答を行わない。

3.2.1 否定形

先行研究 [4]にならない、文末を否定形にすることでデータ拡張を行う。先行研究では与えられた民法条文一覧のみを否定形にすることでデータ拡張して

いたが、提案手法では過去問題の問題文部分も否定形にすることで更なるデータ拡張を行う。

3.2.2 データセットのリスト化

与えられた関連民法条文は、複数の条文に分かれている場合があるため、そのような場合は条文ごとに分割し使用する。分割後の民法条文と、問題文をセットにすることでデータ拡張を行う。

3.2.3 置き替え

民法条文の中には、条番号で他の条文を参照する場合があります。そのような場合、参照先の条文の情報が無いと適切な解答が行えないと考え、人手で条番号を実際の条文の内容に置き替えたデータを作成した。また、これにより文章が長くなりすぎ解答に不利になってしまう可能性も考慮し、置き替え前のデータと置き換え後のデータの両方を使用する。

3.2.4 組み合わせ

分割により民法条文部分に必要なデータが含まれなくなる可能性が考えられるため、分割後のデータから全ての組み合わせで再度民法条文部分を構築し、データ拡張と過不足のない民法条文の作成を同時に行う。

3.2.5 AB 置き換え

過去問題の中には、人物名がアルファベットで登場する場合があります。そのような問題は解答率が低い傾向にある。そこで、問題に二回以上人物名が登場した場合、それらをアルファベットに置き替えることで訓練データを拡張し、このような問題に対応できる可能性を向上させた。

3.2.6 解答の決定

推論の際に、分割された複数の問題が異なる答えを示す場合があります。そのような場合は、Yes と No の解答数を比べ、解答数の多いものをそのモデルの最終的な解答とする。また、結果のばらつきをなくすため同一の訓練データでモデルを5回作成し推論を行う。その際、5個の出力のYesの数とNoの数を再度比較し、多いほうを最終的な解答として出力する。

3.3 アンサンブル

ルールベースのシステムと機械学習ベースのシステムのアンサンブルを行い、精度の向上を図る。ア

ンサンブルの基準にはルールベースのシステムの上位モジュールを使用し、上位モジュールで解答できた問題はルールベースのシステムの解答結果を、それ以外の問題は機械学習ベースのシステムの解答結果を用いる。どこまで上位モジュールとするかを検討するため、全てのパターンでテストを行う。

4 結果

4.1 ルールベースのシステムの結果

COLIEE 2021 に提出した、ルールベースのシステムの実行結果について記述する。提出時のシステムの H18~R01 の過去の問題、R02 のテスト問題に対する結果は表 1 のようになった。なお、ベースラインは今回の改良の行方前の、前年度までのルールベースのシステムである。

表 1 ルールベースのシステム問題実行結果

解答年度	#	ベース ライン	二重構文解析/ 項探索/ 項探索(+条文全体)/ 項探索(+問題全体)/ 項探索(+条文、問題全体)/ (KIS1)		二重構文解析/ 項探索/ 項探索(+条文全体)/ 項探索(+条文全体)/ (KIS2)		項探索/ 項探索(+条文全体)/ 項探索(+条文全体)/ (KIS3)	
H18	36	0.556	0.583	0.583	0.583	0.583		
H19	37	0.514	0.568	0.568	0.568	0.541		
H20	41	0.634	0.659	0.659	0.659	0.61		
H21	54	0.667	0.722	0.722	0.704	0.704		
H22	47	0.553	0.596	0.574	0.553	0.553		
H23	41	0.659	0.61	0.61	0.659	0.659		
H24	79	0.57	0.595	0.595	0.582	0.582		
H25	60	0.583	0.667	0.65	0.633	0.633		
H26	74	0.595	0.635	0.622	0.608	0.608		
H27	49	0.51	0.531	0.531	0.531	0.531		
H28	49	0.633	0.673	0.673	0.673	0.673		
H29	58	0.534	0.569	0.552	0.517	0.517		
H30	70	0.614	0.614	0.6	0.6	0.6		
R01	111	0.622	0.649	0.649	0.64	0.64		
H18~R01 合計	806	0.592	0.623	0.617	0.605	0.605		
R02	81	0.506	0.543	0.531	0.543	0.543		

4.2 機械学習ベースのシステムの結果

機械学習ベースのシステムで過去の問題、テスト問題に対し、5回ずつモデルの作成、推論を行いその結果を統合したものが表 2 である。

表 2 機械学習ベースのシステム問題実行結果

設定					過去問題 (H18~R01) 実行結果	テスト問題 (R02) 実行結果
否定	リス ト化	置き 替え	組み 合わせ	AB 置き 換え		
0	0	0	0	0	458/755 (0.607)	50/75 (0.667)
0	1	1	1	1	476/806 (0.590)	-
1	1	0	1	1	479/806 (0.594)	-
1	1	1	0	1	495/806 (0.614)	-
1	1	1	1	0	494/806 (0.613)	-
1	1	1	1	1	506/806 (0.628)	49/81 (0.605)

4.3 アンサンブルの結果

過去問題,テスト問題の解答についてのアンサンブルを行う. アンサンブルに用いる結果は, ルールベースのシステムは表 1 に記載した KIS1 の結果, 機械学習ベースのシステムは表 2 の最下段に記載した全ての提案手法を用いる設定での実行結果を用いる. なお, COLIEE 2021 提出後にシステムの軽微な不具合を発見したため, 少し修正を加えた. そのためルールベースのシステムの解答結果が少しだけ異なっている.

結果を表に示したものが表 3 である. 例として, 曖昧探索の行に記載されている結果は, ルールベースのシステムの詳細探索, 曖昧探索モジュールまでで解答できた問題についてはルールベースのシステムの解答結果を使用し, 解答できなかったその他の問題については機械学習ベースのシステムの解答結果を使用する.

表 3 アンサンブルの結果

モジュール	節セット	過去問題 (H18~R01) 実行結果	テスト問題 (R02) 実行結果
機械学習ベース	-	506/806 (0.628)	49/81 (0.605)
詳細探索	従来	521/806 (0.646)	51/81 (0.630)
曖昧探索	従来	519/806 (0.644)	52/81 (0.642)
項探 索	文末	従来	504/806 (0.625)
		+AR	46/81 (0.568)
	条文全体	従来	510/806 (0.633)
		+AR	46/81 (0.568)
	問題全体	従来	508/806 (0.630)
		+AR	46/81 (0.568)
	条文、問題全体	従来	508/806 (0.630)
		+AR	43/81 (0.531)
	全探索 (全てルールベースの解答)	従来	503/806 (0.624)
			43/81 (0.531)

5 結果・考察

5.1 ルールベースの結果・考察

過去問題 806 問に対し実行した結果, 表 1 からシステムの改良前と比べ 3%ほど精度が向上した. テスト問題に対して実行した場合, 訓練データと比べあまり精度が高くなかった. しかし, 改良前のルールベースのシステムであるベースラインの正答率は 0.513 と改良後の KIS1~3 より低くなっていた. このことから, 改良により精度が低下したのではなく, ルールベースのシステムが今回のテスト問題と相性が悪かったと考えられる.

5.2 機械学習ベースの結論・考察

表 2 から提案手法を用いたシステムは用いなかったシステムより過去問題の解答精度が向上していた. また, どれか一つの提案手法用いなかったものは, 提案手法を全て用いたものの精度と同等もしくは低下していた. よって全ての提案手法がマイナスには働いていないと考えられる.

テスト問題では提案手法を用いないものの方が精度は高かった. しかし, 提案手法を用いないものの精度が過去問題と比べかなり高いこと, テスト問題の問題数の少なさから, 提案手法の有効性を否定できる結果ではないと考え, 有効性については今後さらなる検討が必要である.

5.3 アンサンブルの結論・考察

表 3 から, 詳細探索モジュールのみ, または詳細探索+曖昧探索モジュールのルールベースのシステムの解答と, それ以外の問題の機械学習ベースのシステムの解答をアンサンブルしたものは, どちらか一方のシステムの解答のみの場合よりも精度が向上した. このことから, アンサンブルに使用するルールベースのシステムのモジュール選択が適切であれば, アンサンブルは非常に有効であると考えられる.

一部の問題のみしか解答できなくとも, 高精度で解答を行えるモジュール, システムがアンサンブルの際には有効であり, そのようなシステムを作成しアンサンブルすることで更なる精度向上が期待できる.

6 おわりに

本論文では, ルールベースのシステムの改良, 機械学習ベースのシステムの実装と改良を行い, それらのアンサンブルを行った. 結果, ルールベースのシステム改良, アンサンブルで特に精度の向上を確認できた.

アンサンブルによる精度の向上は, 精度の高い一つのシステムだけでなく複数のシステムを使用することの有効性を示す結果となった. また, ルールベースのシステムを用いたアンサンブルにより精度が向上したことから, ルールベースのシステムは説明可能な出力といった面のみならず, 精度向上にも貢献できることを示せた.

今後は説明可能な出力を考慮しながら, 機械学習ベースのシステムの改良の有効性の検討, アンサンブルを前提としたモジュール, システムの実装に取り組んでいきたい.

参考文献

- [1] Summary of the Competition on Legal Information Extraction/Entailment (COLIEE) 2021. Juliano Rabelo , Randy Goebel , Yoshinobu Kano , Mi-Young Kim , Masaharu Yoshioka , Ken Satoh. Proceedings of the Eighth International Competition on Legal Information Extraction/Entailment (COLIEE 2021) , 2021 , pp. 1-7.
- [2] Predicate's Argument Resolver and Entity Abstraction for Legal Question Answering: KIS teams at COLIEE 2021 shared task. Masaki Fujita, Naoki Kiyota , Yoshinobu Kano. Proceedings of the Eighth International Competition on Legal Information Extraction/Entailment (COLIEE 2021) , 2021 , pp. 15-24.
- [3] BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Jacob Devlin , Ming-Wei Chang , Kenton Lee , Kristina Toutanova. arXiv:1810.04805v2 , 2019.
- [4] BERT-based Ensemble Methods with Data Augmentation for Legal Textual Entailment in COLIEE Statute Law Task. Masaharu Yoshioka , Yasuhiro Aoki , Youta Suzuki. ICAIL '21: Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law, 2021 , pp. 278-284.
- [5] ParaLaw Nets - Cross-lingual Sentence-level Pretraining for Legal Text Processing. Ha-Thanh Nguyen , Vu Tran , Phuong Minh Nguyen , Thi-Hai-Yen Vuong , Quan Minh Bui , Chau Minh Nguyen , Binh Tran Dang , Minh Le Nguyen , Ken Satoh. Proceedings of the Eighth International Competition on Legal Information Extraction/Entailment (COLIEE 2021) , 2021 , pp. 54-59.
- [6] BM25 and Transformer-based Legal Information Extraction and Entailment. Mi-Young Kim, Juliano Rabelo , Randy Goebel. Proceedings of the Eighth International Competition on Legal Information Extraction/Entailment (COLIEE 2021) , 2021 , pp. 25-30.
- [7] A Pentapus Grapples with Legal Reasoning. Frank Schilder , Dhivya Chinnappa , Kanika Madan , Jinane Harmouche , Andrew Vold , Hiroko Bretz , John Hudzina. Proceedings of the Eighth International Competition on Legal Information Extraction/Entailment(COLIEE 2021) , 2021 , pp. 60-68.
- [8] Using Contextual Word Embeddings and Graph Embeddings for Legal Textual Entailment Classification. Sabine Wehnert , Shipra Dureja , Libin Kutty , Viju Sudhi , Ernesto W. De Luca. Proceedings of the Eighth International Competition on Legal Information Extraction/Entailment (COLIEE 2021) , 2021 , pp. 69-77.
- [9] Question Answering System for Legal Bar Examination using Predicate Argument Structures focusing on Exceptions. Reina Hoshino, Naoki Kiyota , Yoshinobu Kano. Proceedings of the Sixth International Competition on Legal Information Extraction/Entailment (COLIEE 2019) , 2019 , pp. 38-44.
- [10] 日本語構文・格・照応解析システム KNP. 京都大学 大学院情報学研究科 知能情報学専攻 黒橋・楯・村脇研究室. (オンライン) (引用日: 2022 年 1 月 6 日.) <https://nlp.ist.i.kyoto-u.ac.jp/?KNP>.

付録 A

A1 先行研究のシステムのモジュール概要

ルールベースのシステムの先行研究として星野のシステムについて関連研究で述べた。星野のシステムのモジュール構成図は図 A 1 のようになっている。

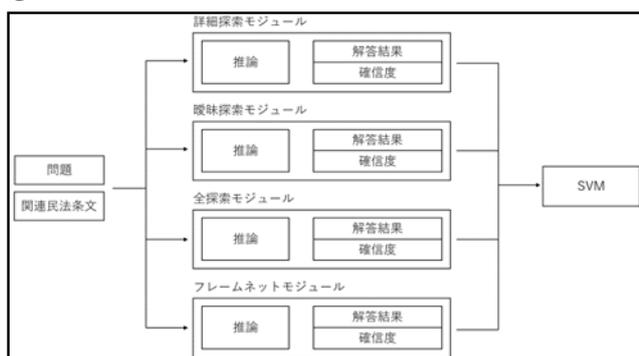


図 A 1 星野のルールベースのシステムモジュール構成図

詳細探索モジュールでは、関連民法条文と問題文の文末の節を比較し、主語、述語、目的語からなる節セットが全て一致していた場合のみ比較を行う。曖昧探索モジュールでは、主語と述語のペア、又は述語と目的語のペアが一致していたときに比較を行う。全探索モジュールでは節セットの比較をせずに推論を行う。

A2 COLIEE 2021 提出結果

ルールベースのシステムを COLIEE 2021 に提出した際の、参加チームの結果一覧が表 A 1 である。KIS1~3 は表 1 と一致している。

表 A 1 COLIEE 2021 Task 4 各チーム提出結果

Team	BaseLine	sid	Correct No 43/All 81	Accuracy
HUKB	HUKB-2		57	0.7037
HUKB	HUKB-1		55	0.6790
HUKB	HUKB-3		55	0.6790
UA	UA_parser		54	0.6667
JNLP	JNLPEncs5C15050		51	0.6296
JNLP	JNLPEncs5C15050SilverE2E10		51	0.6296
JNLP	JNLPEncsBest		51	0.6296
OVGU	OVGU_run3		48	0.5926
TR	TR-Ensemble		48	0.5926
TR	TR-MTE		48	0.5926
OVGU	OVGU_run2		45	0.5556
KIS	KIS1		44	0.5432
KIS	KIS3		44	0.5432
UA	UA_1st		44	0.5432
KIS	KIS2		43	0.5309
UA	UA_dl		43	0.5309
TR	TR_Electra		41	0.5062
OVGU	OVGU_run1		36	0.4444