

判決書自動要約のための修辞役割分類

山田寛章[†] Simone Teufel[‡]† 徳永健伸[†]

[†]東京工業大学 [‡]University of Cambridge

yamada.h.ax@m.titech.ac.jp simone.teufel@cl.cam.ac.uk take@c.titech.ac.jp

1 はじめに

情報処理技術の発達により電子的に参照可能な文書の数は爆発的に増加した。このような状況は法的文書においても例外ではない。裁判官、検事、弁護士等、法律の運用に携わる人々は、日々、過去の関連事件の調査に膨大な時間を費やしている。特に、裁判の記録として最も重要な判決書は事件によっては数十ページを超える長さにも及ぶことも多く、内包される文も日常的に用いられる文よりも長く複雑であり、専門家であっても分析に時間を要する。また、裁判員裁判の導入を端緒とする司法の民主化の流れにも関わらず、裁判所により公的に提供される裁判例検索システム¹は利便性が低く、検索の手がかりとなる要旨も一部文書に付与されているに留まり、利用者が過去の判決書に容易にアクセスできるとは言いがたい。

我々は、日本国の判決書に対する情報アクセスの容易化・効率化を目指し、検索の手掛かりとなる判決の要約を機械的に生成することを目指している。判決書にはその判決結果に至るまでの詳細な議論が記録されており、その議論・論理構造には一定の共通性がある。これは、執筆を担当する者が高度に教育された裁判官であることや、裁判の流れ・裁判制度の構造に起因する記述内容の規則性²によるものである。このような判決書に内包される共通の議論構造を適切に利用できれば、極めて高い品質の自動要約を実現することが可能となる。本研究では、判決書に内包される議論構造情報を要約に応用することを目指し、判決書からの議論構造抽出を目的とする。

2 関連研究

近年、文書中の議論的テキストを抽出・分類することを試みる、議論マイニングと呼ばれる研究が盛んで

ある。Mochales ら [4] は European Court of Human Rights (ECHR) の文書および Araucaria corpus を用いて法的テキスト中からの議論的テキストの検出を行っている。また、Stab ら [6] はインターネット上のフォーラムから抽出したエッセイに対して議論的テキストについてのアノテーション (Claim, Premise 等) を行い、その抽出・分類を行っている。

法律テキスト要約の研究では、科学研究論文の要約のための手法として提案された修辞役割分類分析 [7] を元に法的文書向けに改良して抽出型要約を行う研究が存在する。Hachey ら [3] は英国の判例法システム下における判決書を対象として、文書中の各文に対する修辞役割分類を機械学習を用いて行い、各文が判決書中で果たす修辞役割 (例: 結論, 法律の引用等) を弁別した上で要約生成を行った。また、修辞役割と同様のアプローチを利用して Farzindar ら [2] らがカナダの司法制度下における要約生成、Saravanan ら [5] がインド司法制度下における要約生成の研究を行っている。

本研究と同様に、日本語かつ日本国の法制度下における判決書に対する要約を行っている研究も存在する。阪野ら [13] は、重要文を他の文と弁別するというアプローチをとり、抽出型要約として取り組んでいる。以上の先行研究はいずれも、文書中に含まれる議論構造については考慮していない。

3 議論構造注釈つき判決書コーパス

判決書には一定の共通した議論構造が存在し、これを機械的に抽出することができれば自動要約の品質向上に資する情報を得ることができる。そこで、判決書の議論構造を分析し、日本国判決書に含まれる議論構造情報を定義したアノテーション枠組みを提案した。この枠組みは4種のアノテーションから成る [8]。

¹<http://www.courts.go.jp/>

²判決書を起案する際の指針となる手引き書も存在する [11]。

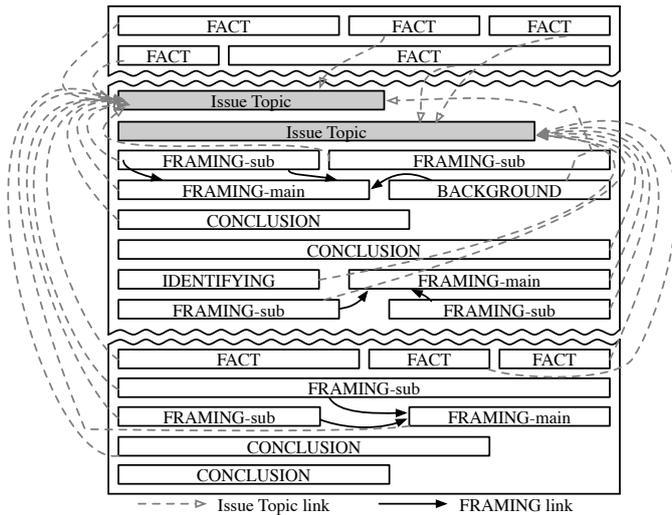


図 1: アノテーションの例

表 1: 日本国判決書のための修辞役割 [8]

ラベル	説明
FACT	証拠や証言、状況やイベントを明示して法廷の場に明示的に事実として提示している部分。
FRAMING-main	CONCLUSION を導くための立論を構成する部分であって、CONCLUSION を直接支える事実認定や主張を行っている部分。
FRAMING-sub	CONCLUSION を導くための立論を構成する部分であって、FRAMING-main を支持する間接事実の認定を行う等、その根拠や理由を示している部分。
CONCLUSION	主要事実の認定など、争点の対象となっているトピックについての結論、または裁判全体の結論を述べている部分。
IDENTIFYING	議論対象となる話題・問題を特定している部分。
BACKGROUND	過去の裁判例や法条を直接引用または参照して、適用している部分。
OTHER	上記のいずれにも適合しなかった部分。

1. 修辞役割分類

判決書からの基本的な議論構造を表現するにあたって、先行研究 [3, 5] に従って修辞役割分析手法を元に表 1 に示す修辞役割を設計した。文書中の各文に対して最も適切なラベルを一つ割り当て³。

2. 争点特定

判決書における議論の中心的话题を示す「争点」

³アノテーション時には、実際には文毎ではなく、句読点を区切りとする単位として修辞構造の特定を行った。その後の分析で単一文内で複数の異なるラベルが付与される割合が 3% 以下であったこと、また、文単位/句読点による単位の双方について分類精度の違いを検討する予備実験を行った結果、文単位による分類器がより高性能であったことから、本稿では文単位でのラベル付与問題として定義している。

ラベル	文数
FACT	4,572
FRAMING-main	3,882
FRAMING-sub	2,339
CONCLUSION	761
IDENTIFYING	451
Issue Topic	127
BACKGROUND	91
OTHER	10,506

表 2: 各カテゴリーの分布

が記述されている文を特定する。

3. 争点関連付け

特定した「争点」と関連する文書中の各文を Issue Topic リンクによって関連付ける。

4. FRAMING リンク

修辞役割分類で FRAMING-main と FRAMING-sub または BACKGROUND として特定されたもの同士の間を FRAMING リンクによって関連付ける、議論上の支持関係を表現する。

以上 4 つのアノテーションにより、判決書の争点を軸とする議論構造を図 1 に示すように階層的に表現することができる。

本稿では、このアノテーションスキームを用いて作成した日本国判決書に対する議論構造注釈付きコーパスを議論構造自動抽出実験のための学習データとして用いる。本コーパスの対象文書は、裁判所 Web ページからダウンロード可能な、下級審裁判所により起案された民事事件裁判例とした。実際にアノテーションされた文書は、平成 15 年 4 月 15 日～平成 28 年 12 月 31 日の間に公開された裁判例のうち、人手による判決要旨が付与されたものを利用した⁴。コーパスは 49 文書、22,729 文から構成される。アノテーションによって付与された各修辞役割の出現数を表 2 に示す⁵。

4 要約デザイン

前節で述べた 4 種のアノテーションにより表現される構造情報を利用することで、元文書そのままの議論構造を保持しながら判決の最重要部分とそれに至る議論の要所を包含できる要約デザインを提案した [9]。

⁴コーパス中には、2 人の異なる作業員によって注釈付けされた 8 文書が含まれる。

⁵表 2 では、便宜上、争点特定で付与される Issue Topic タグも同一表中で集計している。なお、Issue Topic タグと修辞役割タグは排他である。

Type A: 結論部分と重要な議論的テキストからのみなる一般的な重要文抽出型の要約である。修辞役割分類の結果、CONCLUSION と FRAMING-main をラベル付けされた文を重要文とみなし、抽出することによって生成できる。

Type B1: 争点特定, 争点関連付け (Issue Topic リンク) によって表現される議論構造を利用することで、要約中の各文がどの争点について述べているものなのかを関連付ける。これにより、判決書中に複数の争点が混在しているケースでも、各争点ごとに構造化された要約が生成できる。

Type B2: Type B1 を拡張し、特定の争点に焦点を当てた要約である。最も重要な争点について結論だけに留まらず、それを支える議論テキストや判例引用、あるいは事実なども含めて、より粒度の細かい情報を提供する。このため、FRAMING リンクによって表現されるより深い階層の議論構造を利用することが必要となる。

5 修辞役割分類の自動化

本稿では4種のアノテーションのうち、修辞役割分類⁶の自動化を試みる。先行研究 [3, 5] に対応する素性を基本として、日本語を考慮した素性を加えたSVMを用いて多クラス分類として定式化する。言語及び法制度が異なる先行研究が用いた素性が日本国判決書における修辞役割分類に対しても有効なのか、また、機能語表現やモダリティ等の素性が有効なのかを調べ、修辞役割分類の自動化に向けた実験を行う。

5.1 実験設定

実験データとして3節で示した49文書からなるコーパスを使用し、5分割交差検定を行う。SVMはChangら [1] の実装を用いた。また、素性としては、形態素 bigram, 文長, 文の文書中での位置, モダリティ表現, 機能表現, 手掛かり句, 法律の名称を用いた。

このうち、モダリティ表現については、益岡ら [10] の分類に準拠し、主に判断のモダリティ、評価のモダリティとして出現する表現を中心として使用した。ま

た、機能表現は松吉ら [12] の作成した日本語機能表現辞書「つつじ」を使用した。

本実験では、基本的な素性 (形態素 bigram, 文長, 文の文書中での位置) のみを使用したベースラインモデルと、先述の全ての素性を利用した提案手法モデルの2種類のモデルを利用した。いずれのモデルにおいても線形カーネルを用いた。

5.2 実験結果

モデル毎のF値では、ベースラインモデルが0.47、提案手法モデルが0.48となった。この2モデル間で有意水準を0.01としたマクネマー検定を行ったところ、統計的優位差があることを確認できた。しかし、ラベル毎 (表3) では、ベースラインモデルを上回るF値で統計的有意差が確認できたのはOTHERを除けばFRAMING-subのみに留まる。他のラベルに関しては、提案手法ではモダリティ表現や手掛かり句等の素性を導入しているにも関わらず、形態素 bigram を中心とするベースラインモデルとの間に修辞役割分類性能の差がなかったといえる。これは、提案手法で新たに追加された素性によって表現される情報は、基本的には特定の表現をトリガーとする字句ベースのものとなっており、既に形態素 bigram による素性によって十分表現できていたことに起因すると考えられる。

また、モデルに関わらずFACTの識別性能は非常に高い一方で、要約時に重要な要素となるCONCLUSIONはF値で0.25、再現率でも0.31と低い値に留まっている。また、混同行列 (表4) によればFRAMING-subとFRAMING-main, CONCLUSIONとFRAMING-mainの間での誤分類が多発している。元来、CONCLUSION, FRAMING-main, FRAMING-subは議論的テキストとして、文書内の位置的にも表層的な表現としても共通する部分が多く、本実験で使用したような表層表現を単純に使用する素性のみでは十分な分類性能に到達することが難しいと考えられる。また、Issue Topic カテゴリは他のカテゴリと比較して分類性能が極端に低い。本来、Issue Topic自体は修辞役割分類とは別の基準によるラベル設定であり、その抽出のためには本実験で使用した素性群とは異なる素性を利用するなど、別のアプローチが必要であると考えられる。

⁶表1に示される修辞役割に加えて、Issue Topicも分類対象とする。本来はIssue Topicラベルは争点特定という別のタスクに属するものであるが、文に対するラベル付与という点が共通するため、修辞役割分類で同時に抽出することを試みる。

⁷マクネマー検定を0.01の有意水準で行い、提案手法とベースラインモデルの間に統計的有意差が確認できたラベルに対して「*」を付与して表す。

表 3: ラベル別実験結果

ラベル ⁷		精度		再現率		F 値	
		ベースライン	提案手法	ベースライン	提案手法	ベースライン	提案手法
FACT	(FCT)	0.70	0.70	0.83	0.86	0.76	0.77
FRAMING-main	(FRm)	0.57	0.57	0.65	0.63	0.61	0.60
FRAMING-sub*	(FRs)	0.47	0.45	0.22	0.27	0.30	0.34
CONCLUSION	(CCL)	0.51	0.48	0.16	0.17	0.24	0.26
IDENTIFYING*	(IDT)	0.88	0.86	0.65	0.61	0.74	0.71
Issue Topic	(IST)	1.00	1.00	0.01	0.02	0.02	0.03
BACKGROUND	(BGD)	0.31	0.31	0.09	0.09	0.14	0.14
OTHER*	(OTR)	0.92	0.96	0.98	0.98	0.95	0.97
macro avg.*		0.67	0.67	0.45	0.45	0.47	0.48

表 4: 修辞役割自動分類結果の混同行列 (提案手法)

Answer	Prediction								Total
	BGD	CCL	FCT	FRm	FRs	IDT	IST	OTR	
BGD	8	1	21	37	23	0	0	1	91
CCL	1	133	67	477	25	2	0	56	761
FCT	2	22	3954	259	190	11	0	134	4572
FRm	8	64	747	2446	518	15	0	84	3882
FRs	7	6	679	947	638	13	0	49	2339
IDT	0	3	86	48	26	273	0	15	451
IST	0	0	13	6	2	2	2	102	127
OTR	0	47	96	41	7	1	0	10314	10506
Total	26	276	5663	4261	1429	317	2	10755	22729

6 おわりに

本稿では判決書に対する議論マイニングのうち、修辞役割自動分類の実験を行った。今回使用した素性及び分類器は比較的単純なものを使用していたため、特に CONCLUSION, FRAMING-main, FRAMING-sub の議論的テキスト内の分類性能は低いものとなった。しかし、議論的テキストと非議論的テキスト (FACT, IDENTIFYING, OTHER 等) 間では比較的高い F 値を達成しており、議論的テキストの弁別自体には基本的な素性のみでも最低限の分類が行えるという知見が得られた。今後は、議論的テキスト内の分類性能を高めるため、周辺文脈の情報や各文の段落内での位置等のレイアウト情報を素性として導入することなどを検討している。

また、修辞役割分類と同時抽出ができなかった争点特定に加え、争点関連付け及び FRAMING リンク付与についても併せて自動化を行う。

参考文献

- [1] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 2, No. 3, p. 27, 2011.
- [2] Atefeh Farzindar and Guy Lapalme. LetSum, an automatic Legal Text Summarizing system. *Jurix*, pp. 11–18, 2004.
- [3] Ben Hachey and Claire Grover. Extractive summarisation of legal texts. *Artificial Intelligence and Law*, Vol. 14, No. 4, pp. 305–345, 2006.
- [4] Raquel Mochales and Marie Francine Moens. Argumentation mining. *Artificial Intelligence and Law*, Vol. 19, No. 1, pp. 1–22, 2011.
- [5] M. Saravanan and B. Ravindran. Identification of Rhetorical Roles for Segmentation and Summarization of a Legal Judgment. *Artificial Intelligence and Law*, Vol. 18, No. 1, pp. 45–76, 2010.
- [6] Christian Stab and Iryna Gurevych. Parsing argumentation structures in persuasive essays. *Computational Linguistics*, Vol. 43, No. 3, pp. 619–659, 2017.
- [7] Simone Teufel and Marc Moens. Summarizing scientific articles: Experiments with relevance and rhetorical status. *Computational Linguistics*, Vol. 28, No. 4, pp. 409–445, 2002.
- [8] Hiroaki Yamada, Simone Teufel, and Takenobu Tokunaga. Annotation of argument structure in Japanese legal documents. In *Proceedings of the 4th Workshop on Argument Mining*, pp. 22–31. Association for Computational Linguistics, 2017.
- [9] Hiroaki Yamada, Simone Teufel, and Takenobu Tokunaga. Designing an annotation scheme for summarizing Japanese judgment documents. In *2017 9th International Conference on Knowledge and Systems Engineering (KSE)*, pp. 275–280, 2017.
- [10] 益岡隆志. 日本語モダリティ探究. くろしお出版, 2007.
- [11] 司法研修所. 民事判決起案の手引. 法曹会, 10 訂, 2006.
- [12] 松吉俊, 佐藤理史, 宇津呂武仁. 日本語機能表現辞書の編纂. 自然言語処理, Vol. 14, No. 5, pp. 123–146, 2007.
- [13] 阪野慎司, 松原茂樹, 吉川正俊. 機械学習に基づく判決文の重要箇所特定. pp. 1075–1078. 言語処理学会, 2006.