

BERT を用いた刑事裁判例における犯罪構成要件要素の抽出手法について

中村 恒太[†] 阿部 太一[†] 西村 剛輝[†] 中田 己悠[†] 福岡 咲歩[†]
須子 統太[†] 野村 亮[†] 仲道 祐樹[†] 松澤 伸[†]

早稲田大学[†]

1 はじめに

刑事裁判例研究では法律専門家が過去の裁判例を読み込み、各犯罪の成立要件（構成要件）に該当する部分を抽出、比較することで基礎的な分析が行われる。現状はこの作業を人手で行っており、分析する裁判例が多い場合、コストが非常に大きい。

近年自然言語処理を用いた裁判例分析の研究が行われている^{[1][2]}。本研究では刑事裁判例に注目し、自然言語処理を用いて刑事裁判例から構成要件該当部分を自動で抽出するシステムを提案する。

2 刑事裁判例における構成要件要素

刑法の条文は複数の構成要件要素で構成されている。例えば刑法 108 条の現住建造物等放火であれば「放火」、「現住性・現在性」、「焼損」の3つで構成される。構成要件要素とは犯罪が成立するための要件を指し、すべての構成要件要素を満たすことで犯罪が成立する。

刑事裁判では被告人の行為が構成要件要素を満たすか否かが検討される。刑事裁判例は過去の刑事裁判において構成要件該当性がいかなる事実について認められるかが記載されている。

裁判例における構成要件要素の該当部分の抽出は、過去の大量の刑事裁判例を用いた研究に有用で、例えば裁判官の思考の可視化や実際の処罰範囲の理解の一助となる。

3 構成要件要素抽出アルゴリズム

3.1 提案アルゴリズムの概要

図1に提案システムの処理フローを示す。システムには「中心ベクトルの作成」、「分析裁判例の前処理」、「構成要件要素の抽出」の大きく分けて3つの処理がある。

3.2 中心ベクトルの作成

ここでは各構成要件要素に該当する文章を抽出する基準となる中心ベクトルを裁判例から作成する。

本研究では過去の裁判例として『犯罪事実記載の実務』^[3]を用いる。本書は代表的な裁判例を用いて「罪となるべき事実」の実際の記載方法を示したものであり、捜査機関の実務手引書としての機能を有

している。この書に記載されている裁判例に対して法律専門家によって各構成要件要素の該当文章にラベル付けを行い、中心ベクトルを作成した。

本研究ではBERT^[4]によって文章のベクトル化を行う。なお、BERTには東北大学乾・鈴木研究室による日本語学習モデル^[5]を利用した。

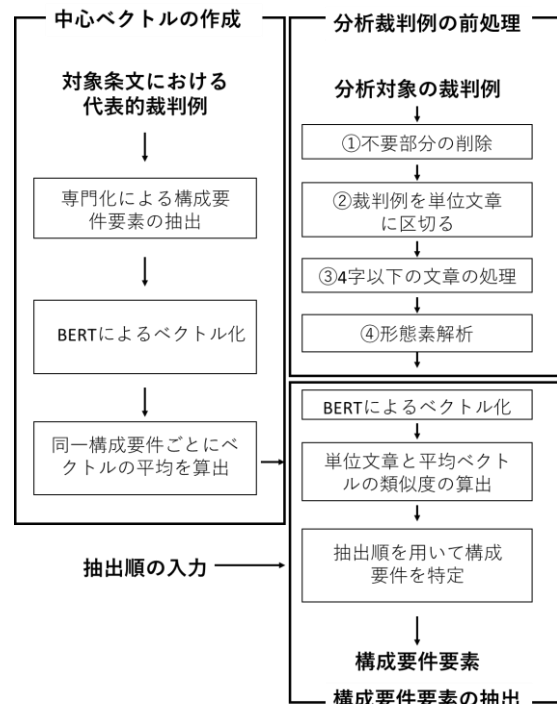


図1: システムの処理フロー

3.3 分析裁判例の前処理

ここでは構成要件要素の抽出対象である新規の刑事裁判例に対して前処理を行う。

図1の①「不要部分の削除」では不要単語の削除と「もって」が入る文章、及びそれ以降の文章の削除を行う。前者について、裁判例中で複数の事件が記載された際に「第一」、「第二」等の単語が用いられる。後者について、通常、罪となるべき事実の文末付近には「もって」の文言とともに条文へのあてはめが行われる。両者ともに構成要件要素に該当しないため削除を行う。

図1の②「裁判例を単位文章に区切る」では裁判例を読点で区切る。刑事裁判例の罪となるべき事実は通常一文で書かれており、句点は最後の一度しか使用されない。そこで本システムでは読点で区切

A Method for Extracting Crime Elements in Criminal Cases Using BERT

[†]Kota Nakamura, Taichi Abe, Goki Nishimura, Miyu Nakata, Sakiho Hukuoka, Tota Suko, Ryo Nomura, Yuki Nakamichi, Shin Matsuzawa, Waseda University

た文章を単位文章として扱う。

図1の③「4字以下の文章の処理」について、裁判例を読点で区切る際、「よって」、「上記Bは」等の単体では本来の意味を持たない単位で区切られる場合がある。そこで本システムでは4字以下の文章は一つ後ろの文章と合成する。

図1の④「形態素解析」では読点区切りの文章ごとに形態素解析を行う。本研究では Sentence Piece^[6]を形態素解析に利用した。

3.4 構成要件要素の抽出

ここでは単位文章と中心ベクトルとの類似度を算出し、類似度に応じて構成要件要素を抽出する。本システムでは同一の文章を複数の構成要件要素の該当部分として抽出しない。重複を避けるために各構成要件要素に番号を割り当て、入力された抽出順で抽出する。

一番目に抽出する構成要件要素に関しては、最も類似度が高い文章を抽出し、二番目以降の構成要件要素に関しては、それ以前に抽出された文章以外で、最も類似度の高い文章を抽出する。

4 評価実験

4.1 実験方法

本システムの評価を刑法108条の現住建造物等放火と刑法246条1項の詐欺罪で行う。評価用データとして、各条文に関連する刑事裁判例を前者について20件、後者について40件を選択し、法律専門家によって各構成要件要素の該当文章にラベル付けを行い、正解データを作成した。

本稿ではすべての抽出順のパターンについて実験を行い、正解率を計算した。

なお、本実験では抽出した文章が正解データの文章に内包されていれば正解とする。各構成要件要素の正解数をデータ数で割ったものを各構成要件要素の正解率、各抽出順での正解数を構成要件要素数とデータ数の積の値で割ったものを各抽出順の正解率とする。

4.2 実験結果

実験の結果を表1,2に記す。

表の1列目の数字は抽出順を表す。例えば、刑法108条の現住建造物等放火であれば「0:放火」、「1:現住性・現在性」、「2:焼損」を割り当てる。抽出順が「102」の場合は、「現住性・現在性」、「放火」、「焼損」の順に抽出する。詐欺罪の構成要件要素も同様である。

5 考察

抽出順ごとの正解率に注目すると両者ともに最も高い正解率が約6割となった。これは実際のシステム運用の際には十分な正解率とは言えない。正解率

表1: 現住建造物等放火における結果

現住建造物等放火 抽出順/構成要件要素	正解率			
	0:放火	1:現住性	2:焼損	検索順ごと
012	0.571	0.667	0.619	0.619
021	0.571	0.667	0.619	0.619
102	0.571	0.667	0.619	0.619
120	0.429	0.667	0.619	0.571
201	0.476	0.619	0.619	0.571
210	0.429	0.667	0.619	0.571

表2: 詐欺罪における結果

詐欺罪 抽出順/構成要件要素	正解率			
	0:欺罔	1:錯誤	2:財物交付	検索順ごと
012	0.415	0.610	0.610	0.545
021	0.415	0.634	0.634	0.561
102	0.512	0.756	0.634	0.634
120	0.537	0.756	0.561	0.618
201	0.488	0.585	0.561	0.545
210	0.537	0.707	0.561	0.602

が約6割にとどまった原因として、中心ベクトルの作成方法が挙げられる。本システムでは各ベクトルの平均をとることにより中心ベクトルを作成しているが、平均化の作業によって特徴的な表現の情報を適切に残すことが出来ていない可能性がある。

4 まとめ

本稿ではBERTを用いた刑事裁判例における犯罪構成要件要素の自動抽出手法について、提案と評価実験を行った。実際にシステムを作成する際には、抽出順を事前に入力する必要がある。しかし、現段階ではその手法までは検討できていない。今後は抽出順の決定手法と精度の向上に取り組む。

参考文献

- [1] 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”, 言語処理学会年次大会発表論文集, No.12, pp. 1075-1078 (2006).
- [2] 野田 奏, 竹内 和広, 三島 聡 “刑事裁判例からの量刑関係情報抽出の試み”, 研究報告人文科学とコンピュータ (CH), No. 2011-CH-91, 1号, pp. 1-6 (2011).
- [3] 末永秀夫・絹川信博・坂井靖・大仲土和・長野哲生・室井和弘・中村芳生, 『犯罪事実記載の実務. 刑法犯』, 実務法規 (2018).
- [4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, n Proc. of NAACL-HLT, pp. 4171-4186 (2019).
- [5] Masatoshi Suzuki, Ryo Takahashi, Pretrained Japanese BERT models, <https://github.com/cl-tohoku/bert-japanese> (2021).
- [6] Taku Kudo and John Richardson. 2018. Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. CoRR, abs/1808.06226(2018).