E-024

機械学習モデル及び BERT による刑種の推定に関する性能比較
Performance comparison of machine learning and BERT for estimating the type of penalty

徳田 翔† 白川 智弘‡ 佐藤 浩†
Sho Tokuda Tomohiro Shirakawa Hiroshi Sato

1. はじめに

今日の司法における裁判では十分な時間をかけた議論が行われ，それに基づいた判決が下されている．このシステムは被告人の権利を保つため必要なプロセスではあるものの，多くの時間を費やしてしまう．現代社会において数多の事例を処理する法曹実務者にとってこの時間を削減することは有意義であると考えられる．

本研究では，法的事例に対してその判断を予測することで，法曹実務者にとってその実務の状況判断する際の一助とすべく，現在までの日本国内の刑事訴訟事例における裁判例を用いて刑種を予測することができないかを検討した．また，認定事実は自然言語によって記述されるため，コンピュータによる自然言語処理が必要とされる．本研究では，近年注目されている，自己注意機構を用いた深層学習を利用した自然言語処理モデルを用いる．本モデルにより，入力された文章から分類器が注目している箇所を抽出することができる．

また，地方裁判所の刑事訴訟事例を対象に，提案手法と既存の機械学習手法による刑種の予測に関する性能比較実験を実施した．その結果，提案手法は既存手法に比べ高い精度での分類が可能であることを確認した．

2. 関連研究

法解釈や裁判の判決予測について機械学習やコンピュータサイエンスの観点から考案されている例[1]として，裁判官の思考過程に基づき，裁判のプロセスに基づいて結果を推論する方法について考案しているものがある．

判例・判決文を元に，自然言語処理技術によって判決を予測したものとしては欧州人権裁判所の判決文から特定の条文における事例に対して違反か非違反かの予測が約 79%の精度でできることを示すもの[2]がある．

以上のように，法律解釈の手法の提案や海外における判決予測等は行われているものの，国内の司法判断の予測については未だ手を付けられてはいないという現状がある．

3. データ及び方法

3.1 データ

日本の刑事訴訟事例を扱うため，判例のデータベースである D1-law.com 第一法規総合法情報データベース[3]から判例のテキストデータを収集しデータセットを作成した．

データセットに使用した裁判例は事例の事実が必ず記載されている地方裁判所の刑事訴訟事例にかかる第 1 審判決とし，出力される結論を単一のものとするため被告人が一人のものを用いた．

裁判例の主な構造を表 1 に，裁判例の一例を図 1 に示す．このうち，事実から刑種を特定するため，刑罰について記載されている 1. 主文とその犯罪事実が記載されている 2. 事実をそれぞれ抽出し，主文の刑種に応じてラベル付けを実施した．

表 1　裁判例の構造

項目	内容
主文	判決の内容
事実	犯罪事実、証拠、前科など
裁判所の判断	審理による法的判断
法令の適用	判断に基づき適用される法条
量刑の理由	主文の判決に至った理由

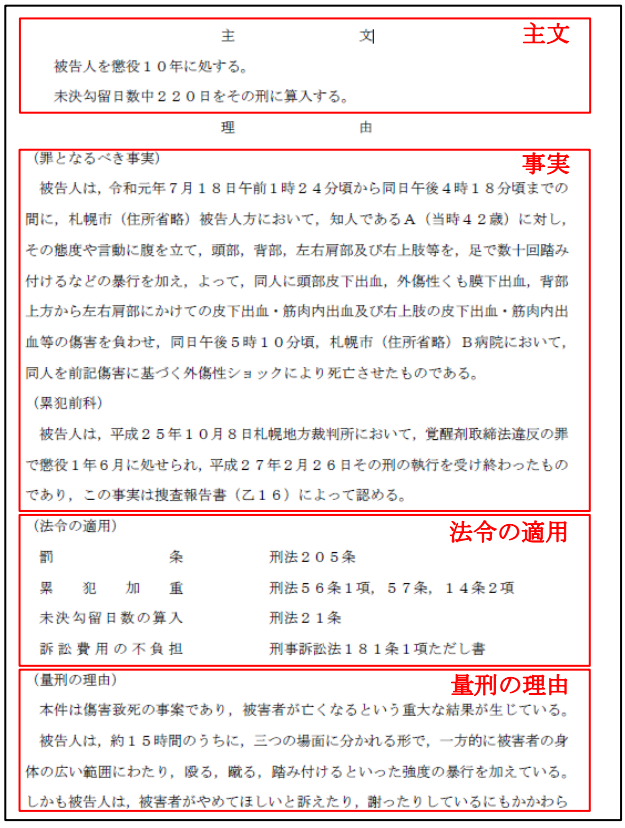
主　　文　　主文
被告人を懲役１０年に処する。
未決勾留日数中２２０日をその刑に算入する。
理　　由
（罪となるべき事実）　事実
被告人は，令和元年７月１８日午前１時２４分頃から同日午後４時１８分頃までの間に，札幌市（住所省略）被告人方において，知人であるＡ（当時４２歳）に対し，その態度や言動に腹を立て，頭部，背部，左右肩部及び右上肢等を，足で数十回踏み付けるなどの暴行を加え，よって，同人に頭部皮下出血，外傷性くも膜下出血，背部上方から左右肩部にかけての皮下出血・筋肉内出血及び右上肢の皮下出血・筋肉内出血等の傷害を負わせ，同日午後５時１０分頃，札幌市（住所省略）Ｂ病院において，同人を前記傷害に基づく外傷性ショックにより死亡させたものである。
（累犯前科）
被告人は，平成２５年１０月８日札幌地方裁判所において，覚醒剤取締法違反の罪で懲役１年６月に処せられ，平成２７年２月２６日その刑の執行を受け終わったものであり，この事実は捜査報告書（乙１６）によって認める。
（法令の適用）　法令の適用
罰条　刑法２０５条
累犯加重　刑法５６条１項，５７条，１４条２項
未決勾留日数の算入　刑法２１条
訴訟費用の不負担　刑事訴訟法１８１条１項ただし書
（量刑の理由）　量刑の理由
本件は傷害致死の事案であり，被害者が亡くなるという重大な結果が生じている。
被告人は，約１５時間のうちに，三つの場面に分かれる形で，一方的に被害者の身体の広い範囲にわたり，殴る，蹴る，踏み付けるといった強度の暴行を加えている。しかも被告人は，被害者がやめてほしいと訴えたり，謝ったりしているにもかかわら

図 2　裁判例の一例

†防衛大学校理工学研究科 Grad. School of Math. and Comp. Sci., National Defense Academy
‡長岡技術科学大学大学院工学研究科 Grad. School of Engineering, Nagaoka University of Technology

3.2 方法

刑種の推定には BERT[4] (Bidirectional Encoder Representations from Transformers) を利用した．これは Transformer[5] のエンコーダーを利用したモデルで，事前学習によってパラメータ調整されたモデルにタスクに応じたレイヤーによってファインチューニングすることで様々な自然言語タスクを処理することができるモデルである．

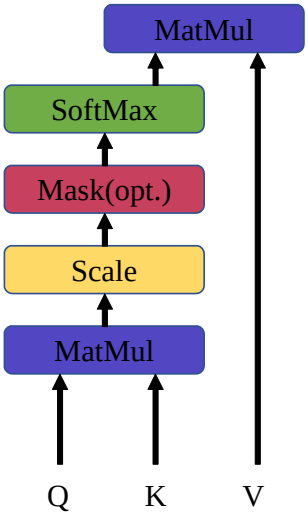

図 2　Transformer に使われている Self-Attention[5]

この BERT 及び Transformer には，Self-Attention 機構というものが用いられている．これは入力データの単語間の関連度を示す仕組みであり，入力データを検索用の Query，その対象となる Memory として二つの入力データを用意，Memory を Key と Value に分離した辞書型オブジェクトとし，Key と Query の内積を求めこれに対応する Value を取り出すものをいう（図 2）．

なお，Transformer の用いられる Attention は以下の式 (1) によって求められる．式内の d_k はスケーリング因子である．

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (1)$$

4. 実験

作成したデータセットにより，以下の二つの実験を実施した．

1　BERT を用いた文書分類

2　先行研究[2] に基づく機械学習を用いた文書分類

なお，BERT へ入力できる文章の最大長が 512 トークンであり，また，モデルの性質上 2 つのスペシャルトークンを解析対象の文章に加える必要があるため，作成したデータセットのうち 510 トークン以下のデータのみを使用した．その際，データの偏りをなくすためラベルごとのデータ数をそろえている．データセットのラベルごとのデータ数を表 1 に示す．

表 2　ラベルの種類及びデータ数

	無期懲役 LIW	懲役 IW	禁錮 I	罰金 FI	懲役・罰金 IW F	総数
件数	54	54	54	54	54	270

4.1 BERT を用いた文書分類

BERT の文書分類モデル (BertForSequenceClassification[6]) によって文書分類を実施した．これは BERT の最終層に分類用の全結合層を加えたものであり，その出力によって文書の分類をするモデルである．

4.2 機械学習を用いた文書分類

先行研究[2] に基づく実験では，Bag-of-Words によりデータセットの入力文章を単語の出現頻度によりベクトル化し，それを SVM によって分類している．本研究でも同様の手法を用いた．ただし，先行研究においては，違反又は非違反の 2 値分類であるが，当実験では多値分類である点が異なる．

今回の実験では全入力データの単語出現頻度を計算し，文章中によく出る付属語を除くため、上位 500～2500 位を用いた 2000 次元のベクトル表現と、名詞・動詞・形容詞・形容動詞を抽出した，上位 2000 位までの単語を用いた 2000 次元のベクトル表現を作成し，2 種類の実験を実施した（図 3）．

データセットの全文章から形態素解析により単語とその頻度を抽出

20	分	370
21	甲	363
22	5	363
23	事実	362
24	同人	361
25	六	355

回数カウントベクトル（2000次元）
Case1={0, 1, 0, 5,…,0}
Case2={2, 0, 4, 1,…,4}
…
Case270={1, 0, 3, 0,…,1}

ベクトルを正規化
Case1={0, 0.05, 0, 0.25,…,0}
Case2={0.02, 0, 0.04, 0.005,…,0.04}
…
Case270={0.1, 0, 0.3, 0,…,0.1}

図 3　Bag-of-Words による ベクトル化のイメージ

先行研究[2]において用いられたモデルは SVM であったが，他の分類のための機械学習モデルにおいても検証を実施し，その精度を確認した．なお，分類に使用したモデルは下記の通りとした．

・SVM
・重回帰分析
・線形 SVM
・決定木分析
・ランダムフォレスト
・K 近傍法
・ガウシアンナイーブベイズ
・勾配ブースティング法
・SGD

5. 考察

5.1 結果

実験結果について表 3，4 及び 5 に示す．表 3 は各モデルのラベルごとの F 値を示したものである．機械学習手法によるものは左側に全単語を利用した場合の結果，右側に自立語のみを利用した結果を表記している．なお，最下段はモデルの正解率 (Accuracy) を示している．

表 4 に BERT モデルのラベルごとの正解数（混同行列）を示し，表 5 にラベルごとの再現率，適合率及び F 値を示す．

5.2 考察

5.2.1 モデルごとの評価

モデルごとの正解率を見ると，提案手法である BERT を用いたものが正解率 0.75 と最も高いスコアを記録した．その他の機械学習による分類では，正解率を見ると全単語を利用したものは線形 SVM でその値は 0.67，自立語のみのものはランダムフォレスト及び勾配ブースティングによるものが最も高い正解率であり，その値は 0.69 であった．先行研究手法で用いられた SVM では，正解率が全単語利用のものは 0.66，自立語のみのものは 0.65 であり，全体の中では高めのスコアであった．

5.2.2 ラベルごとの評価

ラベルごとの F 値を見ると，懲役，罰金のスコアがとりわけ低く記録された．表 4 及び 5 によれば，この二つのラベルで相互に誤答しているものが多いことがわかる．

これは同じ犯罪行為でも懲役と罰金の両方が規定されており，状況などによりこの判断が変わってしまうためと考えられる．誤答した事例の一例として，図 4 に示す．この事例は住居侵入罪についてのものであるが，住居侵入罪は懲役・罰金どちらも規定しており，ここで示す例ではその分類で誤っている．

表 3　各モデルの F 値と精度

	BERT	**SVM**		LR		L-SVM		決定木	
		全単語	自立語	全単語	自立語	全単語	自立語	全単語	自立語
LIW	0.91	**0.76**	**0.76**	0.57	0.52	0.74	0.64	0.57	0.71
IW	0.47	**0.48**	**0.39**	0.38	0.09	0.43	0.23	0.44	0.36
I	0.91	**0.75**	**0.81**	0.63	0.45	0.76	0.71	0.50	0.81
FI	0.57	**0.63**	**0.59**	0.46	0.47	0.63	0.52	0.47	0.44
IW F	0.96	**0.69**	**0.73**	0.62	0.70	0.71	0.73	0.56	0.71
ACC	0.75	**0.66**	**0.65**	0.53	0.49	0.67	0.59	0.50	0.60

	RF		K近傍		GNB		勾配		SGD	
	全単語	自立語	全単語	自立語	全単語	自立語	全単語	自立語	全単語	自立語
LIW	0.69	0.81	0.49	0.70	0.61	0.63	0.61	0.85	0.77	0.54
IW	0.46	0.40	0.21	0.28	0.39	0.25	0.41	0.47	0.43	0.03
I	0.76	0.87	0.44	0.68	0.69	0.72	0.62	0.88	0.58	0.53
FI	0.51	0.60	0.04	0.22	0.54	0.44	0.44	0.50	0.55	0.30
IW F	0.56	0.77	0.10	0.76	0.67	0.69	0.70	0.76	0.65	0.73
ACC	0.58	0.69	0.35	0.57	0.58	0.55	0.54	0.69	0.60	0.48

正解カテゴリ: 罰金
予測カテゴリ: 懲役
(犯罪事実)被告人は、正当な理由がないのに、令和元年 10 月 26 日午前 1 時 59 分頃、(住所略)の A らが居住する 3 階及び 4 階に通じる外階段並びに 4 階通路に、同外階段の 2 階から 3 階に通じる途中に設置された鉄製片開き戸(高さ約 2 メートル)を乗り越えて侵入した。
(証拠の標目)(括弧内の甲乙の数字は、検察官請求の証拠番号を示す。)

図 4　罰金の事例を懲役と誤答した例

表 4　BERT モデルによる刑種予測結果（混同行列）

		予測				
		LIW	IW	I	FI	IW F
正解	LIW	15	2	0	0	0
	IW	1	7	0	3	0
	I	0	1	10	0	0
	FI	0	8	1	8	0
	IW F	0	1	0	0	11

表 5　BERT モデルによる刑種予測の評価値

	適合率	再現率	F値
LIW	0.94	0.88	0.91
IW	0.37	0.64	0.47
I	0.91	0.91	0.91
FI	0.73	0.47	0.57
IW F	1.00	0.92	0.96
正解率			0.75

6. 結論

本研究では，自然言語処理モデルである BERT を用いて，日本国内の刑事訴訟事例の刑種の分類を行うことで，犯罪事実からある程度の精度で判決を予測することができることを示し，また，その精度は従来手法によるものよりも優れていることが分かった．

今後の課題として，本手法によるモデルでは入力データが最大 512 トークンであり，長い文章が入力できない点を解決する必要がある．犯罪事実の文章はその内容が複雑になればなるほど長くなる傾向があり，実務上の需要が高いと考えられる．

また，懲役と罰金のように 2 種類以上規定されている犯罪について詳細に判断するため，言語モデルに対して法律知識を付与する必要があると考えられる．

最後に，懲役や罰金等一部の刑種は期間や金額といった量刑を決定しなければ刑罰の予測にはならない．そのため，これらの刑種を予測した際には量刑まで決定できるモデルを考案する必要があると考えられる．

参考文献

[1] 太田 勝造, “ AI 裁判支援システムへの人々の期待と受容”, 人工知能学会全国大会論文集，33 号 (2019).
[2] Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preoţiuc-Pietro, Vasileios Lampos, “Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspec-tive”, PeerJ Computer Science 2:e93 (2016).
[3] https://dtp-cm.d1-law.com/ (Accessed in Jun.11, 2021)
[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, arXiv preprint arXiv:1810.04805 (2018).
[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, “Attention is all you need”, Advances in Neural Information Processing Systems，6000/6010 (2017).
[6] https://huggingface.co/ (Accessed in Jun.11, 2021)