

令和元年度 情報工学コース卒業研究報告要旨

外山 研究室	氏 名	吉 岡 大 貴
卒業研究題目	汎用事前学習モデルを用いた 重要文抽出による法律の要約	
<p>法令は社会の設計図であり，法令を理解することによって社会制度や権利・義務などの国民の生活に関する多くの情報を得ることができる．しかし，法令は長い文や構文が複雑な文で構成されるため，一般に読みにくく理解しにくいとされる．また，細かく規定する条文などがあるため文数が多くなるという点も法令の読みにくさ，理解しにくさの原因となる．そこで，法令の理解を促進させるために法令の自動要約が研究されている．</p> <p>一般に，文書の自動要約について，当初はルールベースによる手法が提案されてきたが必ずしも十分な性能が得られず，統計ベースによる手法が提案されるようになった．統計ベースによる手法には学習コーパスが必要である．従来の研究では人手により学習コーパスを作成してきたが，コストが大きく，大量のコーパスを得られなかった．</p> <p>先行研究の一つである佐藤らの論文（以下，先行研究）では，政府が作成する法令の要約である「法令のあらまし」（以下，あらまし）に着目し，従来より多量のコーパスを自動で作成した．また先行研究では，あらまし文に含まれる文を重要文として抽出することにより要約を行った．その際，重要分抽出を分類問題として定式化し，ルールベースの既存手法に対する優位性を示した．加えて，ランダムフォレスト分類器がSVMや決定木などの他の分類器よりも有効であることを示した．</p> <p>本研究では，分類器による重要分抽出によって法律の要約を行うアイデアを踏襲し，分類器として汎用事前学習モデルBERTを用いることを提案する．BERTは，Attentionのみを用いたニューラル言語モデルであるTransformerからエンコーダ部分を使用し，事前学習タスクを工夫することにより双方向学習を実現して高い性能を出したモデルである．また，出力結果の後に目的のデータを取得するための層を追加で接続することでさまざまなタスクに簡単に対応させることができるという特徴がある．さらに，ニューラル言語モデルの中でも，BERTは大量の一般文による事前学習を行うため，学習コーパスの少ないタスクでも利用することができる．実際，山腰らはBERTが日本の法令に関するタスクで有効であることを示しており，今回のタスクにおいてもBERTが高い性能を発揮することが期待される．</p> <p>本研究で学習と評価に用いるコーパスは，先行研究で法律文とあらまし文の自動対応付けから作成したものを使用した．学習コーパスは1,055法律，24,500文（正例8,918文，負例15,582文）からなり，テストコーパスは学習コーパスとは別の33法律，1,121文（正例789文，負例332文）からなる．</p> <p>評価実験の結果，提案手法は抽出性能において再現率85.49%，精度87.67%，F値84.49%を示し，全ての項目で先行研究の手法よりも高い性能を示した．特に精度については約6ポイントの向上がみられ，有意水準5%でランダムフォレストより優れていた．また，要約において重視される再現率についても約1ポイントの向上がみられた．しかし，要約の自動評価指標であるROUGEスコアについては，既存手法に僅かに劣る結果となった．特に，ROUGE-1のスコアについては約1ポイント劣っていた．この原因として，今回のタスクでは要約文の長さに制限がないため，再現率が同程度である場合，精度が高いほど抽出割合が低くなることが挙げられる．BERTは重要な文だけを精度よく抽出する傾向があり，それに伴ってROUGEスコアが低くなってしまったと考えられる．今回の手法は抽出性能の面で向上がみられ，提案手法の有効性を示すことができた．</p>		