

論文の内容の要旨

専攻名 システム創成工学専攻

氏名 金澤 政和

日本国憲法第14条には「すべて国民は、法の下に平等であつて、人種、信条、性別、社会的身分又は門地により、政治的、経済的又は社会的関係において、差別されない」と記載されているが、現実には、地理的な制約、法律に関する知識の多寡により、十分な平等が保証されているとは言い難いのが現実である。この理念を実現するためには、司法へのアクセス（Access to the Justice）の平等性を実現することが不可欠である。近年のインターネットの普及により、この理念を実現する可能性は高まっていると考えられる。法へアクセスの平等を実現するためには、アクセシビリティ、ユーザビリティ、リライアビリティが実現されなければならないが、司法にICTを導入することで、これを実現しようという研究が各国で行われている。

ICT化した裁判制度をサイバーコートと呼ぶ。司法へのアクセスの平等を実現するためには、申立に始まる裁判手続きのオンライン化、判決の公開などがある。しかし、実際には判決の公開率は最高裁判決文で0.9%程度と極端に低い。判決文はプライバシー保護の観点から特定組織や人名、地名等は匿名化（伏字化）することが必要条件であるが、この作業を手で行っている関係で処理が追いつかないという実態によるものである。そこで本論文では、全判決の公開を目指し、ニューラルネットワークを利用した言語処理による伏字化処理の自動化を行うことを目的とする。具体的には、過去に、手作業で匿名化された判例を真のデータとして学習データに用い、機械学習することで、未だ匿名化（伏字化）されていない判例を入力すると自動的に伏字化するシステムの構築を目指す。文章の中から特定の固有名詞を抽出するという固有表現抽出技術の研究分野では、近年、BERT（Bidirectional Encoder Representations from Transformers）と呼ばれる新しい自然言語処理技術が出現し90%の正解率という高い成果を収めている。ただし、事前に大量の学習データを用意する必要があり現実の判例に適用するのは容易ではない。そこで本論文では自然言語処理の分野で広く利用されているニューラルネットワークで判例文書を学習させて伏字対象となる単語を検出する方法について研究した。第1段階として、Bi-directional LSTM（Long Short-Term Memory）モデルを利用して実験した結果、一般名詞と姓名の誤判定などが発生し、伏字検出精度が上がらないことが判明した。このため、第2段階として、Bi-directional LSTMに与える入力データに品詞情報（POS）タグを付加したモデルを提案した。また、これに加え、前処理の改善を行った。その結果、伏せ字の検出精度を、飛躍的に向上させることができた。本モデルを実際の判決文に適用し評価を行った結果、実験した判例の約90%で、実際の伏字の中からどれだけ伏字と判定したかの割合を示す再現率が70%以上であった。このことから、伏字検出システムとして実用化の可能性を示したと結論づけた。

以下に、本論文の構成と各章の内容を述べる。

第1章では、本研究の背景・目的と概要を述べる。背景として、2017年に内閣府が発表した未来投資戦略の中で挙げられた「裁判のIT化」について解説し、その中の課題の一つである判決文の自動匿名化を実現することが重要であることを述べ、本研究の目的と意義を明らかにする。

第2章では、サイバーコートを構成する「e提出(e-Filing)」「e事件管理(e-Case Management)」「e法廷(e-Court)」について解説する。

第3章では、サイバーコートの先駆けとなったアメリカ、シンガポール、韓国などの主要国の裁判のIT化状況を述べる。

第4章では、日本のサイバーコートの実現に向けた取り組みの現状を解説し、裁判のIT化の必要性および重要性について述べる。

第5章では、日本のサイバーコートを実現させるための課題である、システムの構築、セキュリティ、判決文等の公開について解説し、本研究テーマとなっている伏字処理の必要性と現状について述べる。

第6章では、サイバーコート実現における課題の一つである裁判の公開化を実現するため、日本の判決文中の匿名語を検出するためのニューラルネットワークによる先行研究について述べる。さらに、先行研究における実験結果と問題点について概説する。

第7章では、先行研究の精度を向上させるため、ニューラルネットワークに品詞情報(POS)タグを付加したモデルを提案する。実験結果では、提案モデルの精度が向上したことを示し、更に前処理アルゴリズムの改善によって先行研究結果に比べて伏字予測確率は88%の精度向上したことを述べる。次に実用化の可能性を見出すため、実判例文を本モデルに適用し、得られた課題に対して前処理を改善した結果、テストした判例数の90%で再現率が70%以上の値が得られ実用化の可能性のあることを述べる。

第8章では、本モデルの更なる精度向上策を述べるとともに、これまでの実験結果を踏まえた実用化に向けた今後の課題を提示する。

第9章では、本研究を総括し、得られた成果をまとめる。さらに、目標である判例の全件公開化へ向けた今後の課題・展望について述べる。