

論文審査の結果の要旨

専攻名 システム創成工学専攻

氏名 金澤 政和

本論文は、「A study on automatic detection method of the confidential word in the Japanese precedents（日本の判決文における匿名語の自動検出手法に関する研究）」と題し、日本の判決文の公開のために必要となる、判決文の伏字化処理の自動化を行う手法とその実用性について述べたものである。2017年に内閣府が打ち出した「未来投資戦略」の一つに、「裁判の手続等のIT化」が挙げられている。裁判のIT化には、司法へのアクセスの平等化を実現するために判決文の公開が欠かせない要素となっているが、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、この作業を人手で行っているため処理が追いつかないという実態によるものである。このような背景を踏まえ、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手法を提案した。提案手法では、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を構築した。

本論文は全9章から構成されている。第1章では、本研究の背景・目的と概要が述べられている。第2章では、サイバーコートの概念とその効果を概説している。第3章では、サイバーコートの先駆けとなった主要国の裁判のIT化状況を述べている。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と裁判のIT化の主な内容を述べている。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を述べている。第6章では、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方法を述べると共に先行研究内容とその実験結果を概説している。第7章では、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し、実験結果で精度が向上したことを示している。更に前処理アルゴリズムの改善によって伏字化の精度が大幅に改善されたことを述べている。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実用化に向けた今後の課題を提示している。第9章では、本研究を総括している。

本研究で得られた成果は以下のようにまとめられる。

- 1) 学習データに判決の原文を大量に用いることが困難であるため、公開されている伏字化された判決文を学習データとして構築したニューラルネットワークを用いて、原文を伏字化するネットワークを構築できることを示した。これは、ターゲット単語を伏字とすべきかどうかを周辺単語から判断するネットワークを構築したためである。
- 2) 前処理において形態素解析器による単語の分かち書きを行う際、単語の品詞情報を抽出して、ニューラルネットワークに埋め込みベクトルとして付加したニューラルネットワ

ークモデルを新たに提案した。これにより、品詞情報を用いないニューラルネットワークと比べて伏字の予測精度が大きく改善することを示した。

- 3) 学習データとなる伏字化された判決文は、様々な記号が用いられているが、これを統一した記号を用いるよう前処理を改善することにより、更に伏字の予測精度が改善することを示した。
- 4) 提案手法の性能評価において、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%以上であることを示した。また、改善前のモデルと比較すると、再現率が10%から80%に改善する判例もあった。以上より、判決文匿名化の自動化処理実現への第一歩を示した。

本論文については、2021年2月5日にオンライン(Zoom)において、審査委員全員及びこの分野の研究者の出席のもとに公聴会が開催され、研究内容の発表及び質疑応答が行われた。公聴会終了後ただちに学位審査委員会を開催し、本論文の内容について詳細に検討した。その結果、本研究成果は司法分野への学術的寄与が高く、実用分野での貢献も期待できると判断された。また、研究内容の学術的水準と独創性においても極めて優れていると評価された。

よって、本論文は博士（工学）の学位論文に値するものと認める。