

令和 4 年 6 月 22 日現在

機関番号：25408

研究種目：挑戦的研究(萌芽)

研究期間：2017～2021

課題番号：17K18549

研究課題名(和文) 外交データ分析に関する人工知能(AI)の開発に向けて

研究課題名(英文) Towards applying the Artificial Intelligence (AI) system in the studies with diplomatic records

研究代表者

長岡 さくら (NAGAOKA, Sakura)

叡啓大学・ソーシャルシステムデザイン学部・准教授

研究者番号：10550402

交付決定額(研究期間全体)：(直接経費) 4,800,000円

研究成果の概要(和文)：本研究は、将来、人工知能(AI)技術を外交史料研究において利活用するために
行う探査的な研究である。

近年、外交記録公開によって公開される外交文書の量が著しく増大している。このような状況で、公開された
膨大な外交文書を、一人の研究者が、特定の研究目的群に従って、全ての関連ある資料を発見した上で、一定の
期間内に体系的な検討まで行うことは困難を極める。

そこで、本研究では、近年、発展がめざましいAI技術を外交史料研究に利活用するために必要となる手法に
ついて検討を行い、AI技術が、どの程度、外交研究において利活用可能なのか、又、どのように役立つのかに
ついて整理を行った。

研究成果の学術的意義や社会的意義

近年のAI技術の発展はめざましい。しかし、本研究開始以前、外交史料研究の領域においてAI技術の利活
用は模索されてこなかったように見受けられる。それは、AI技術を利用する必然性がなかっただけでなく、本
領域においてどの程度AI技術が利用可能であるのか検証され、その成果が共有されていないことに起因してい
るように思われる。

それ故、本研究によって、外交史料研究においてAI技術が一部利用可能であることを確認できたことは、本
領域におけるAI技術の利用可能性への道を開くものでもあると考える。この点で、外交研究分野におけるAI
技術の利活用可能性を整理したこと自体が学術的意義や社会的意義を有すると考える。

研究成果の概要(英文)： This research is an exploratory study for the application of technology
for building an Artificial Intelligence (AI) system in the studies with diplomatic records.

Currently, vast amount of diplomatic records has been de-classified in Japan by enactment of the
Directive of the Ministry of Foreign Affairs of Japan and Public Records and Archives Management
Act. Under these circumstances, it is extremely difficult for a scholar to discover relevant records
with his or her specific research objectives within the extreme amount of these records and to
examine them systematically within a limited research timeline.

Therefore, in this research, we were exploring the conditions for applying the AI system in the
studies with diplomatic records. In particular, we tried to identify how much extent the technology
for building the AI system could be applied for the studies with diplomatic records and how it could
be worked in such studies.

研究分野：国際法学

キーワード：外交文書 AI技術 データ分析 外交研究 国際法学

1. 研究開始当初の背景

(1) 研究代表者は本研究開始以前、海洋の事例を材料として、国家が一方的に行った行為に関して国際法上どのように評価されるのかについて一貫して検討してきた。とりわけ、博士論文では、ある国家が設定した基線に対して抗議を行う第三国の事象を通じて国際法上の合法性及び正当性について外交史料を用いて検討を行った。その結果、沿岸国が行った基線設定行為が国際法に合致していないとして第三国による抗議を受けた場合、そのほとんどは沿岸国によって無視されていることを明らかにした(長岡さくら、『国際法における直線基線の設定と第三国の対応 - 国家実行における抗議の実効性 - 』九州大学提出博士学位論文(2016年3月))。

(2) 研究代表者は、博士論文の執筆過程において、数千頁にわたる日本の外交文書及び米国の国家実行に関する米国行政文書を検討した。その際、外務大臣訓令の制定(2010年)及び公文書等の管理に関する法律の施行(2011年)以降、外交記録の公開が急加速されつつある現実に直面した。公開された外交文書数が飛躍的に増加し、ビッグデータとなりつつあることは、一面において、様々な事象を体系的に検討するための素材・下地が整うというメリットがある。その反面、個々の研究者が外交史料を丹念に読み込む伝統的手法では、検討対象とする事象について一定期間内に体系的な検討結果を得ることが困難になりつつあるのではないかと思いついた。

(3) このような現状を打破するためには、個々の研究者が行う外交史料の検討プロセスの一部をAI技術によって補完することができれば、ビッグデータとなった外交史料を駆使した研究が更に発展する可能となるのではないかと考え、本研究の構想に至った。

2. 研究の目的

(1) 本研究は、将来、AI技術を外交史料研究において利活用するために行う探求的な研究である。

(2) 従来、法学・政治学分野における外交研究では、伝統的に、個々の研究者が、紙媒体の資料を一つ一つ確認しながら分析する手法が用いられてきたと言えよう。

日本の外交文書は、主として1976年以降の外交記録公開制度によって公開が進んできた。従来はその公開速度が緩やかであったため、伝統的な分析手法のみに依拠したとしても、個々の研究者が一定期間内に体系的な分析を行うことにそれほど困難が生じていた訳ではなかったように思われる。ところが、2010年以降、外務大臣訓令の制定や公文書等の管理に関する法整備が行われたことにより、外交文書の公開が加速され、公開済みの外交文書の量が著しく増大している。

(3) このような状況は、外交文書を用いて研究を行う者から見れば喜ばしい現象である。しかしながらその一方で、次のような懸念もある。即ち、公開された膨大な外交文書を、一人の研究者が、特定の研究目的に従って、全ての関連ある資料を探し出し、読みこなし、分析を行い、それを一定の期間内に体系的な検討として公表することは果して可能なのであろうか？

これに加え、膨大な量の外交史料の分析を行うにあたっては、さらに分析を困難にする問題が存在する。即ち、提供される外交史料の媒体及びデータの種類の種類に起因する問題である。

これまでに外交記録公開制度等によって入手可能な日本の外交史料は、二つの媒体に分類することができる。即ち、紙媒体史料及び画像データ化された史料である。しかし、これまで日本の外交史料の多くは紙媒体によるものであった。また、仮にその紙媒体の外交文書が画像データ化されていたとしても、それは単に紙媒体であった史料が画像化されただけのデータであって、史料として分析を行うためには更にその画像化されたデータについて文字データ化を行うなどの工夫が必要となることが課題であった。

(4) そこで本研究では、近年、発展がめざましいAI技術を外交史料研究に利活用するために必要となる手法について検討を行い、現状におけるAI技術が、どの程度、外交研究において利活用可能なのかについて整理し、利活用の手法について提案を試みることにした。検討を行うにあたっては、研究代表者がこれまでに国際法研究において培った外交文書の分析を、AI技術で補完する際に用いることができる手法について整理し、現存するAI技術で何がどこまでできるかについて検証を行うことにした。また、どのような困難があるかについても、AI技術開発に関わる研究者の助言を受けつつ整理を行うことにした。

3. 研究の方法

(1) 本研究では、研究総括において外交史料研究を補完するAI技術利用あるいは今後のAI開発に向けた提案を行うこと目標とした。その際、AIに期待する能力は、研究代表者がこれまでに行ってきた外交文書を用いての研究活動の一部(あるいは全部)について、AIが補助補完

(あるいは代替)する能力を備えることである。そのためには、これまで研究代表者が博士論文執筆過程において収集し検討を重ねてきた国際法分野の外交史料等をサンプルデータとして用い、現存するA I技術の到達度を明らかにすることが最も適していると考えた。

(2) そこで、本研究では、以下の三段階に分けて検討を行うこととした。

まず、サンプルデータをパソコンでの分析可能な文字データとして認識させるための準備作業を行う。その際、既存の光学文字認識(OCR: Optical Character Reader)ソフトによる文字認識が可能な資料と不可能な資料に分けて作業工程を比較検討し、現在の技術的到達点と課題を整理する[文書認識段階]。次に、文字データ化されたサンプルデータに対し、複数のデータマイニングソフトを用いて解析結果を抽出する。そして既存のA I技術利用状況について整理する[文書解析段階]。最後に、文書認識段階及び文書解析段階から得られた結果を検討し、A I技術の利活用の現状と課題についてについて整理する[A I技術の現状総括段階]。

4. 研究成果

(1) 本研究では、まず、サンプルデータ(文字データ及び画像データ)をパソコンでの分析可能な文字データとして認識させるための準備作業として、画像データ資料に対し、既存の複数のOCRソフトを用いてOCR作業(文字画像の認識)を行い、市販されている汎用的なOCRソフトにおける技術的到達点と課題を整理した。

本研究では、当初、紙媒体で提供され、これを画像化した一連の外交史料を主たるサンプルデータとして用い、研究を進めた。その結果、使用したサンプルデータには、言語の種類、文字の態様、文書の形式、綴り込まれている史料の多様性等の問題から、市販されている汎用的なOCRソフトを用いた簡便な文字データ化には、ソフトの調整だけでは越えられない課題があることが判明した。複数のOCRソフトを試してみたが、一データ中に複数の言語等が混在する外交史料についてOCRソフトでは文字データ化すら十分にこなせないことが判明した。

(2) 次に、市販されている汎用的なOCRソフトを用いての作業(文字画像の認識)の結果として得られた技術的課題を回避するための作業を中心に進めた。

とりわけ、画像化された史料のOCRに関しては、パターン認識による文字認識アルゴリズム自体を構築可能な数値解析ソフトに変更して対応することにした。数値解析ソフトを用いての文字認識アルゴリズムの構築には、プログラミング環境を整備する必要があり、新たに当該環境も整えた。また、文字画像認識に用いるサンプルデータの見直しも行い、書体や言語等の混在しない19世紀に公刊された資料の画像データを利用した。なお念のため、OCRソフトによる文字データ化の作業結果と数値解析ソフトを用いて作成した文字認識アルゴリズムによる文字データ化の作業結果とを比較した。

その結果、数値解析ソフトによる作業結果の方が、OCRソフトによる作業結果よりも高精度であることが確認できた。文字データ化された文書についての解析に関しては、質的データ分析(QDA: Qualitative Data Analysis)ソフトの有用性について確認を行った。

(3) 更に、これらの検討結果を受け、文字データ化された文書の解析を中心に研究を進めた。

文字データ化された文書の解析について、QDAソフトによる解析は、A I技術の開発(実装)についての研究として本研究を展開させるには適していないことが判明していた。そのため、文字データ化された文書の解析作業を文字認識アルゴリズムと一体的に進めることを可能にするため、文字認識アルゴリズムを構築した数値解析ソフト環境を用いて、新たに文書解析のアルゴリズム構築に取り組んだ。

(4) なお、上述の検討と並行して、次の二点について、本研究との関連において法的観点から整理を行った。

まず、近年、様々な国内外の政府機関、国際機関、学術団体等において作成や提言が活発に行われている、「A I」の開発や利活用に関するガイドライン(「A I原則」)に関し、しばしば指摘されている「A I」と人権や倫理との関連性について検討・整理を行った。

又、紙媒体文書の画像認識・文字データ化やデータ化された文書の整理・加工・分析等の作業に伴う当該データ類の著作権の問題についても整理を行った。

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 0件／うち国際共著 0件／うちオープンアクセス 4件）

1. 著者名 長岡 さくら	4. 巻 2019
2. 論文標題 国際法研究と人工知能（AI）技術 - - 外交史料の利活用に向けて - -	5. 発行年 2019年
3. 雑誌名 九州法学会会報	6. 最初と最後の頁 40～43
掲載論文のDOI（デジタルオブジェクト識別子） 10.20661/klia.2019.0_40	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 長岡 さくら	4. 巻 2
2. 論文標題 人工知能（AI）技術の外交史料研究への利活用の探究 - - 紙媒体史料の文字認識と課題 - -	5. 発行年 2020年
3. 雑誌名 福岡工業大学総合研究機構研究所報	6. 最初と最後の頁 155～160
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 長岡 さくら	4. 巻 4
2. 論文標題 人工知能（AI）と人権 - - AI原則の法的意義 - -	5. 発行年 2021年
3. 雑誌名 福岡工業大学総合研究機構研究所報	6. 最初と最後の頁 85-89
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計1件（うち招待講演 0件／うち国際学会 0件）

1. 発表者名 長岡 さくら
2. 発表標題 国際法研究と人工知能（AI）技術 - - 外交史料の利活用に向けて - -
3. 学会等名 九州法学会第124回学術大会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

長岡さくら、「人工知能（AI）技術の外交史料研究への利活用の探究 - 紙媒体史料の文字認識と課題 - 」福岡工業大学環境科学研究所環境研究発表会2019（2019年）
 長岡さくら、「人工知能（AI）技術の国際法研究への利活用に向けて」福岡工業大学環境科学研究所環境研究発表会（2019年度環境科学研究所環境研究発表会）（2019年）
 長岡さくら、「人工知能（AI）と人権 - AI原則の法的意義 - 」第15回福岡工業大学2020年度環境科学研究所研究発表（2021年）
 長岡さくら、「人工知能（AI）を巡る国際法上の論点 - 国際法上の著作権・序論 - 」福岡工業大学総合研究機構環境科学研究所第16回環境研究発表会（2022年）
 中川智治、「国際法分野における人工知能（AI）技術の利活用を巡る法的課題」令和元年度福岡工業大学研究成果報告会（2019年）
 中川智治、「国際法学における位置データを含む外交データ分析の課題について」第15回福岡工業大学2020年度環境科学研究所研究発表（2021年）
 中川智治、「法学分野へのAI技術利用の研究動向について」福岡工業大学総合研究機構環境科学研究所第16回環境研究発表会（2022年）

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	中川 智治 (NAKAGAWA Tomoharu) (50336046)	福岡工業大学・社会環境学部・教授 (37112)	

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
連携研究者	藤岡 寛之 (FUJIOKA Hiroyuki) (10349798)	福岡工業大学・情報工学部・教授 (37112)	
連携研究者	小林 和博 (KOBAYASHI Kazuhiro) (00450677)	東京理科大学・理工学部経営工学科・講師 (32660)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関