

## 論文内容の要旨

博士論文題目

Advancing Rare Disease Research by Unifying Both Supervised and Unsupervised Methods

(教師あり手法と教師なし手法の統合による希少疾患研究の進展)

氏名 YAO Lean Franzl Lim

(論文内容の要旨)

最新の自然言語処理は、近年、大規模言語モデルや生成 AI の発達に伴い劇的な進化を遂げている。しかし、一部の専門領域においては、自然言語処理の解釈可能性や説明可能性が問題になる場合もある。本研究では、希少疾患を題材にして、伝統的な機械学習の関連性と最新の言語処理を比較しながら2つの研究を行った。

まず、学術論文データベースを用いて、世界中の医学研究の動向を定量化し、可視化する研究を行なった。これは、教師ありの言語処理に基づいた。開発したツールは、ステークホルダーが、重要な疾患カテゴリーへの動向を調査するのに役立つ。

次に、ソーシャルメディアデータに教師なし学習を適応することで、話題(トピック)とその変化を抽出し、COVID-19 のコミュニティへの影響を調査した。開発したツールは、どのような言説がコミュニティに広がっているのか洞察を得るために用いることができる。

これらの研究を通して、本論文は、希少疾病研究の領域における自然言語処理について、教師なし、教師ありといった2つの機械学習手法の特性を明らかにするものである。

|    |                     |
|----|---------------------|
| 氏名 | YAO Lean Franzl Lim |
|----|---------------------|

(論文審査結果の要旨)

最新の自然言語処理は、近年、大規模言語モデルや生成 AI の発達に伴い劇的な進化を遂げている。しかし、一部の専門領域においては、自然言語処理の解釈可能性や説明可能性が問題になる場合もある。本研究では、希少疾患を題材にして、伝統的な機械学習の関連性と最新の言語処理を比較しながら2つの研究を行った。1つは、学術論文データベースを用いて、世界中の医学研究の動向を定量化するものであり、最新の教師ありの言語処理手法に基づいている。もう1つは、ソーシャルメディアデータに教師なし学習を適応し、トピックを可視化するものであり、コミュニティの解析など社会学研究に貢献する。これらの2つの研究は、現在の医療情報研究の全体の縮図とも言え、本論文の方法論、考察は、今後の医療情報の研究の方向性を議論する上での有益な材料となる。

研究が細分化、かつ加速しつつある自然言語や医療分野における研究の方法論は多様化し、どのような手法が良いか俯瞰的に考えることが難しい。このような状況の中、実際の医療情報データを扱い、幅広い手法を検討した本研究の意義は高い。よって、博士(工学)の学位論文としての価値があるものと認める。