

自然言語処理技術を用いた 医療記録からの症状抽出方法に関する研究 ～患者の主観的情報に焦点を当てて～

岡山大学病院薬剤部

大野由紀子

Yukiko Ohno

この度は、Postdoctoral Award という名誉ある賞を賜りましたことを誠に光栄に思うとともに今後も本賞を励みに研究と臨床業務に邁進していきたい。本稿では、自身の学位研究に取り組んだきっかけと研究内容について概説する。

研究のきっかけ

私は学部生の際に臨床現場に近い研究に挑戦したいと思い、慶應義塾大学病院薬剤部と連携したAIホスピタルを研究テーマとした。AIホスピタルは、内閣府の戦略的イノベーション創造プログラム「AI(人工知能)ホスピタルによる高度診断・治療システム」に基づいた取り組みである。その中で、私は医薬品の適応症・禁忌をモニタリングするシステムの性能評価に取り組んだ。そのシステムは試験的に開発されたものであり、実用化には至らなかったものの、その研究を通して、今後は臨床現場においてもAIを活用していくことが求められると直感した。そこで、自身が臨床現場とAIをつなぐ役割を担いたいと考え、大学院へ進学した。大学院では、当研究室にて行われた先行研究¹⁾に機械学習手法を取り入れた研究テーマに携わることとなった。

研究背景

自然言語とはプログラミング言語と対照的に私たちが日常で使用する言語を指す。自然言語の

データは構造化されておらず、そのままではコンピュータによる解析ができないため、自然言語処理技術を用いて解析する必要がある。近年は生成AIの登場により、自然言語処理技術の性能が飛躍的に向上しており関心が高まりつつあるが、依然として発展途上にある。

本研究では自然言語処理技術の中でも固有表現認識という技術を用いて医療テキストから疾患名や症状といった情報を収集した。本研究を始めた当初、患者の主観的情報を分析対象とした機械学習の研究は英文や特定の患者層を対象とした研究に限られていた。その中で、和文テキストを対象とし、特定の症状に限定されない汎用性のある機械学習ベースのシステム構築に取り組むこととした。

患者の主観的情報が含まれる情報源として患者の闘病ブログやSOAP形式で記載された医療記録の主観的情報が挙げられる。医療従事者の中でも、薬剤師は薬学的観点から患者指導を行うため、薬剤師記録に有害事象情報が豊富に含まれる。また、主観的な症状は患者が第一に気づくため、S情報を分析対象とした有害事象モニタリングの有用性も期待される。こうした背景から薬剤師記録に含まれる患者の主観的情報を分析対象としたシステムの実現を着想した。

研究成果の概要

本研究は既存システムの活用可能性の検討と、

新規システムの構築および性能向上から構成される。まず既存システムの活用可能性の検討として、医師記録を学習済みの既存システムを薬剤師記録に適用できるか性能評価を行った。さらに、既存システムを有害事象モニタリングへ応用できるか調査した。この取り組みを通して既存システムを薬剤師記録の主観的情報に用いるには課題があることが明らかとなったため、薬剤師記録の主観的情報を教師データとした新規システムの構築に取り組んだ。またシステムの性能向上のために、学習量増加と教師データとする記録の質の二点から検討を行った。

以降、新規システムの学習量増加に伴う性能向上について紹介する。²⁾

教師データとして、2018年度の慶應義塾大学病院の薬剤管理指導記録に含まれる主観的情報6,559症例、12,004件を用いた。データを前処理し、疾患名・症状にアノテーション(正解情報の付与)を行い、正解情報付きのデータを作成した。データを10分割し、10段階にわけて学習量を増やしながら学習に用いた。各学習量について、10分割交差検証法により、主要な機械学習モデルの1つである bidirectional encoder representations from transformers-conditional random field を学習させ、性能を評価した。1,200件から12,004件の学習量増量により、F1-scoreは0.67から0.82へと上昇した。学習量の増加とともに、性能向上の幅は小さくなったものの、大量のデータを学習させることにより、高性能な自然言語処理システムを構築することができた。

今後の展望

本研究においては、臨床情報の収集と限定的な有害事象モニタリングに終始したが、今後は収集

した情報を構造化する技術とその情報をもとにモニタリングを行う仕組みづくりが必要である。本研究において生成AIは用いられていないが、生成AIを活用して機能の拡張やシステムの性能評価を行うことで、早期に本モニタリング機構の実現が期待される。

謝 辞

最後に、日頃よりご指導ご鞭撻をいただきました大谷壽一先生(慶應義塾大学病院薬剤部、慶應義塾大学医学部・薬学部)、望月眞弓先生(慶應義塾大学薬学部)、青森 達先生(高崎健康福祉大学薬学部)、荒牧英治先生(奈良先端科学技術大学院大学)、慶應義塾大学薬学部臨床薬学講座の先生方と関係各位に深く感謝申し上げます。また、新天地にて研究を後押ししていただき、本賞へもご推薦いただきました座間味義人先生(岡山大学病院薬剤部)に厚く御礼申し上げます。

参考文献

- 1) Usui M, Aramaki E, Iwao T, Wakamiya S, Sakamoto T, Mochizuki M, Extraction and Standardization of Patient Complaints from Electronic Medication Histories for Pharmacovigilance: Natural Language Processing Analysis in Japanese, *JMIR Med Inform*, 2018, **6**, e11021. doi: 10.2196/11021.
- 2) Ohno Y, Aomori T, Nishiyama T, Kato R, Fujiki R, Ishikawa H, Kiyomiya K, Isawa M, Mochizuki M, Aramaki E, Ohtani H, Performance Improvement of a Natural Language Processing Tool for Extracting Patient Narratives Related to Medical States From Japanese Pharmaceutical Care Records by Increasing the Amount of Training Data: Natural Language Processing Analysis and Validation Study, *JMIR Med Inform*, 2025, **13**, e68863. doi: 10.2196/68863.