

様 式 C-7-1

令和 2 年度科学研究費助成事業（科学研究費補助金）実績報告書（研究実績報告書）

所属研究機関名称		奈良先端科学技術大学院大学	機関番号	1 4 6 0 3
研究 代表者	部局	先端科学技術研究科		
	職	特別研究員(DC2)		
	氏名	柴田 大作		

1. 研究種目名

特別研究員奨励費

2. 課題番号

20J12069

3. 研究課題名

電子カルテデータからの患者状態表現の自動抽出に関する研究

4. 研究期間

令和 2 年度～令和 2 年度

5. 領域番号・区分

—

6. 研究実績の概要

本研究では、診療録から臨床研究や治験を実施する際に必要な患者情報である疾患の既往歴と生活歴に関する情報を自動で抽出するためのシステムの開発と評価を行った。本研究において抽出する情報は医療者と相談の上で決定し、既往歴は8つの疾患（糖尿病、高血圧、脂質異常症、腎障害、脳梗塞、脳卒中、くも膜下出血、冠動脈疾患）を、生活歴は喫煙歴と飲酒歴に関する情報を対象とした。診療録は基本的にフリーテキストとして記載されるため、表現の揺れが多発する。また、記載される情報は患者が現在罹患している疾患だけではなく、否定されている疾患、過去に罹患した疾患や家族に関する情報などさまざまである。そのため辞書などを用いたルールベースでこれを行うことはコストが大きいと考えられる。そこで本研究では機械学習を用いた既往歴・生活歴の抽出を行った。機械学習の適用には、アノテーションされたデータ（教師データ）が必要であるため、最初にアノテーション定義の検討とそれに基づいた診療録へのアノテーションを実施した。アノテーションでは、診療録を文単位に分割し、各文にそれぞれの疾患と生活歴に関する情報の有無（現在、過去、現在の否定、過去の否定、プロブレムリストやその他など）と構造化に関する情報（現病歴、既往歴、生活歴、家族歴など）に関する情報を付与した。これを診療録285件（10,457文）に対して実施し、実験材料とした。機械学習による識別では、近年の自然言語処理タスクで高い精度を達成しているモデルを用いた実験を行った。診療記録で事前学習されたBERTやWikipediaで事前学習されたBERTなどを用いて性能を比較した結果、診療記録で事前学習されたBERTを用いることで、疾患に関する情報はF1値が平均で0.546、構造化に関する情報はF1値が平均で0.791と一定の精度で識別できることが明らかとなった。

7. キーワード

医療情報学 医療言語処理

8. 現在までの進捗状況

区分

理由

翌年度、交付申請を辞退するため、記入しない。

2 版

9. 今後の研究の推進方策

翌年度、交付申請を辞退するため、記入しない。

10. 研究発表（令和2年度の研究成果）

〔雑誌論文〕 計0件

〔学会発表〕 計0件

〔図書〕 計0件

11. 研究成果による産業財産権の出願・取得状況

計0件（うち出願0件／うち取得0件）

12. 科研費を使用して開催した国際研究集会

計0件

13. 本研究に関連して実施した国際共同研究の実施状況

—

14. 備考

—