

先端科学技術研究科 修士論文要旨

所属研究室 (主指導教員)	ソーシャル・コンピューティング (荒牧 英治 (教授))		
学籍番号	2111187	提出日	令和 6年 1月 22日
学生氏名	西谷 実紘		
論文題目	生成アプローチによる患者表現の標準化		
要旨			
<p>近年、患者の声を医療に生かす試みが盛んに行われている。患者の声を捉えることで主観的な症状の捕捉や電子カルテの入力の簡便化を行うことにつながる。しかし、患者の声を医療に生かすためには、患者の自由な発話を機械処理可能な形に変換する必要がある。そこで本研究では、患者の発話のうち症状表現を標準的な表現に変換するためのリソースである大規模患者表現辞書を構築し、それを利用した機械学習モデルを構築した。大規模患者表現辞書は、クラウドソーシングによるアンケートなどから患者表現とその標準形のペアを大規模に収集することで構築した。収集した患者表現は6,506件であり、そのうち標準病名が付与されているものは6,393件である。大規模患者表現辞書は、Webページ (https://sociocom.naist.jp/patient-dic/)にて公開している。次に、大規模患者表現辞書と臨床現場で使われる病名を収集した万病辞書から複数の設定のデータセットを構築し、生成アプローチを用いて患者表現から標準病名を生成した。実験では比較のため、分類モデルを用いた。その結果、最も精度が高いデータセットにおいて生成モデルが0.653という結果を出し、分類モデルを上回った。また、実験結果から、生成アプローチを用いることで、正解の病名に類似した文字列や正解とは異なるが意味の通る病名を生成できることが明らかになった。これにより、患者が実際に使いそうなテキストを低コストで実現することができる。今後、これを用いた擬似的な実験や教師データの構築が期待される。</p>			