

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	1911402	提出日	令和 3年 7月 28日
学生氏名	氏家 翔吾		
論文題目	Improving the methods for disease entity linking in low-resource settings 少資源下における病名エンティティリンキング手法の改善		
要旨			
<p>医学論文からの情報の抽出は臨床研究や医療現場での応用に向けた重要な課題である。これを実現するためエンティティリンキングシステムが研究されている。エンティティリンキングは、文書から該当する箇所を抜き出す固有表現抽出と、抜き出した文字列を事前に定義された概念に紐づける曖昧性解消により実現される。それぞれのモデルの学習、推論には大量の学習データや大規模な類義語辞書が必要となるが、それらのデータが利用できない場合も多い。本論文では、言語資源が少ない場合に生じる課題解決を目的として、既存手法の拡張をはじめとする各種手法を提案する。まず、既存のエンティティリンキング手法を拡張し、学習データに出現しない概念を扱う手法について述べる。既存研究と比較し、提案手法は既存手法の精度を保ったまま学習データに現れない概念についても一定の精度で予測可能であることを示す。次に、小規模な類義語辞書のみが使用可能な場合にも学習、推論可能な曖昧性解消手法を述べる。類義語辞書の規模を増減させて実験を行い、類義語辞書が小規模なほど提案手法が有用であることを示す。</p>			