The University of Osaka
Institutional Knowledge Archive

Title	BERT-based Transfer Learning in Sentence-level Anatomic Classification of Free-Text Radiology Reports
Author(s)	西垣, 大毅
Citation	大阪大学, 2024, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/96332
rights	
Note	やむを得ない事由があると学位審査研究科が承認したため、全文に代えてその内容の要約を公開しています。全文のご利用をご希望の場合は、大阪大学の博士論文についてをご参照ください。

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

論　文　内　容　の　要　旨
Synopsis of Thesis

氏　　名 Name	西垣　大毅
論文題名 Title	BERT-based Transfer Learning in Sentence-level Anatomic Classification of Free-Text Radiology Reports （BERTを用いた転移学習によるフリーテキストの放射線診断報告書の文単位での解剖学的分類）

論文内容の要旨

〔目　的(Purpose)〕

フリーテキストの放射線診断報告書の文レベルの解剖学的分類における、臨床コーパスで事前学習したbidirectional encoder representations from transformers (BERT)モデルを用いた転移学習の性能を評価する。また、学習データの少ない解剖学的クラスにおける性能も検証する。

〔方法 (Methods)〕

2005年12月から2020年12月までにMedical Imaging Clinic (大阪府豊中市)で全身PET/CT検査を受けた患者の放射線科診断報告書から抽出された6,272文を対象とした。これら各文は、体の部位(脳、頭頸部、胸部、腹部、四肢、脊椎、その他)に応じて、2人の放射線科医の合議によりラベル付けされた。ラベルの分布は不均衡であった。学習データが少ないクラスでは、脳クラスで10文(0.9%, 10/1102)、四肢クラスで18文(1.6%, 18/1102)、脊椎クラスで15文(1.4%, 15/1102)しかなかった。BERTを用いた転移学習アプローチは、2つのベースラインの機械学習モデルであるbidirectional long short-term memory (BiLSTM)およびcount-basedアプローチと比較された。各手法について、適合率再現率曲線下面積(AUPRC)、受信者動作特性曲線下面積(AUC)を計算し、比較した。AUCの統計学的比較は、DeLong検定を用いて行い、$p<0.05$を統計的有意とした。

〔成績(Results)〕

BERTはmacro-averaged AUPRC=0.88の分類性能を達成し、ベースラインを上回った。BERTのAUCは、すべてのクラスでBiLSTMのAUCを上回り、脳・胸部・腹部・その他クラスではcount-based法のAUCを上回った。BERTのAUPRCは、ラベル付きの学習データが少ないクラスでも、ベースラインより優れた性能であった（脳: BERT, 0.95, BiLSTM, 0.11, count-based, 0.41; 四肢: BERT, 0.74, BiLSTM, 0.28, count-based, 0.46; 脊椎: BERT, 0.82, BiLSTM, 0.53, count-based, 0.69）。

〔総　括(Conclusion)〕

フリーテキストの放射線医学報告書の文レベルの解剖学的分類において、BERTモデルを用いた転移学習は、ラベル付き学習データが少ない解剖学的クラスであっても、BiLSTMおよびcount-basedアプローチを上回る性能を示した。

論文審査の結果の要旨及び担当者

（申請者氏名）　西垣　大毅

		（職）	氏　　名
論文審査担当者	主　査	大阪大学特任教授	木戸尚治　署名
	副　査	大阪大学教授	武田理宏　署名
	副　査	大阪大学教授	岡田隨象　署名

論文審査の結果の要旨

放射線診断報告書には重要な臨床情報が含まれており、診断や治療の決定、後方視的研究、放射線画像のアノテーションなど、様々な目的で使用されている。しかしながら、多くの報告書はフリーテキストで記載されているため、情報抽出に多大な労力や時間を要する。したがって、多くの研究で、臨床報告書を自動的に構造化し整理しようとする試みが行われている。

本研究では、報告書の構造化の一環として、文単位での解剖学的分類を試みた。先行研究で用いられた機械学習アプローチでは、モデルの性能が学習データ量に依存するため、学習データが少ない解剖学的クラスでは性能が低かった。我々は、大規模な臨床コーパスで事前学習したbidirectional encoder representations from transformers (BERT) モデルを用いた転移学習を用いることにより、学習データが少ないクラスにおいても、ベースラインの機械学習アプローチよりも高い性能を達成した。また、我々は、BERTモデルが事前学習によって、本研究で用いた学習データのみからよりも多くの医学知識を獲得した可能性を示した。我々の研究結果は、フリーテキストの放射線診断報告書の構造化において、学習データが少ない場合でも、BERTモデルを用いた転移学習が有用である可能性を示すものであり、新規性があるため、学位の授与に値すると考えられる。