

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	2311402	提出日	令和 7年 7月 22日
学生氏名	大槻 優佳		
論文題目	医療テキストの構造化		
要旨			
本研究は、臨床文書に埋もれる高価値情報を低コストかつプライバシー保護を維持したまま構造化することを目的とし、自由記述が多い脳卒中データベース項目と表記揺れが顕著な病理バイオマーカ記述を対象に、Few-shot LLMとData-Augmented SLMを組み合わせた二段階パイプラインを開発した。第1段階では、わずか五～十例の例示で推論する大規模言語モデルにより、脳卒中発症様式や治療介入、転帰評価といった複数ラベルを抽出し、従来のT5ベース細調整モデルと同等のF1スコアを達成した。第2段階では、LLMが生成した多様な擬似病理文を小規模モデルに再学習させるデータ拡張戦略を採用し、バイオマーカ正規化でF1>0.90を達成した。いずれもGPU8 GB環境でリアルタイム処理が可能であり、追加学習用データ共有が難しい医療現場における実運用性を実証した。さらに、タスクを跨いだモデル設計指針と生成データ品質評価プロトコルを提示し、他疾患領域への横展開可能性を示した。本アプローチは、脳卒中急性期診療支援の迅速化と腫瘍バイオマーカ研究の再現性向上を同時に促進し、精密医療の情報基盤整備に寄与する。			