令和７年度厚生労働科学研究費補助金（地域医療基盤開発推進研究事業）
分担研究報告書

【分担研究６】情報の削減を抑えて診療記録を短縮する手法を開発する研究

園部　真也・宮本　太郎・川本　章太・中村　直毅・工藤　憲一

【要旨】
診療記録を LLM に読解させることは情報抽出や文脈判定において有用であるが、診療記録の有する情報量が多すぎて計算効率が悪いという課題がある。ルールベースアルゴリズムを用いた適切な診療記録の情報削減は、運用の精度を維持しつつ計算効率を飛躍的に向上させる。

【目的】
診療記録を LLM に読解させることは情報抽出や文脈判定において有用であるが、診療記録の有する情報量が多すぎて計算資源が追い付かないという課題や、計算に時間がかかりすぎて現実的でないという課題がある。さらに、医療情報である診療記録は、個人情報保護の観点からデータの流出が制限されるため、大規模な計算資源が使用できないという側面も有している。
一方で診療記録においては、同じ情報が複数回にわたり記載されていることや、読解の用途によっては必要のない情報が記載されていることも多々あり、このような診療録を LLM に読解させることは必ずしも効率的な手法とは言えない。
本研究では、このような情報を削減してから LLM に読解させることで、計算を効率化させ、ひいては LLM に診療記録を読解させる仕組みの実現可能性を高めることを目的とする。

【方法】
診療記録をそのまま読解させた場合（原文）の、それに要する計算時間と、その読解に基づいて抽出される情報の正確性を、評価した。また、診療記録を削減させてから読解させた場合（抜粋）の、それに要する計算時間と、その読解に基づいて抽出される情報の正確性を、同様に評価した。これらの評価を 300 名の乳癌患者にて施行した。
診療記録は、最終的に抽出したい情報の関連キーワードを列挙し、それぞれのキーワードを含む時間的に最新の文を抜粋する、というルールベースアルゴリズムにより削減した。
観察研究であり、オプトアウト方式で実施した。

【結果】
原文は平均で約 120,000 字から構成されていた。抜粋は平均で約 1,000 字から構成されていた。原文から抜粋を作成する時間は、積算しても無視できるほど短い時間であった。抜粋の読解に要する時間は原文の読解に要する時間の 1%を下回っていた。読解に基づいて抽出される情報の正確性は、抜粋では原文よりもやや低い結果であったが、許容できる範囲であった。

【考察】
ルールベースアルゴリズムを用いた診療記録の情報削減と LLM による読解の組み合わせは、膨大な情報を有する非構造化データである診療記録の利活用において必須の技術である。情報削減のアルゴリズムをさらに工夫することで、より高く精度を維持したまま情報削減ができると考えられる。

【結論】
適切な診療記録の情報削減は、運用の精度を維持しつつ計算効率を飛躍的に向上させる。

【論文発表】
該当なし

【学会発表】
園部真也，宮本太郎，川本章太，井汲沙織，中村直毅，藤井進．臨床研究の対象となる候補患者を医療記録に基づいて抽出するシステムの構築．第 45 回医療情報学連合大会．

【知的財産権の出願・登録状況】
該当なし