

Utilization of Healthcare Information Extracted from Medical Documents and Social Media Texts

氏 名 西山智弘

研究室名 ソーシャル・コンピューティング研究室

主指導教員名（論文博士の場合は推薦教員名） 荒牧 英治

内容梗概（1ページ目に収めること）

近年、自然言語処理（NLP）の進展により、電子カルテ（EHR）やソーシャルメディアなど、非構造化医療データから有用な情報を抽出する技術が実用化の段階に近づいている。本論文では、こうしたデータを臨床現場で実用的に活用するための方法論を探求し、情報の抽出、構造化、適用における課題を解決することを目指す。ヘルスケア分野において正確で包括的なデータの利用は、医療の質と安全性を向上させるために不可欠である。そこで、本研究では非構造化データである医療文書とソーシャルメディアの2種類のテキストからの情報抽出とその利用について実用化の方法を検討する。医療文書については、固有表現抽出（NER）技術を用いた情報抽出と構造化の方法を検討し、従来の構造化データでは得られなかった症状や副作用の可視化を可能にする手法を開発する。ソーシャルメディアデータからの情報抽出については、ソーシャルメディアデータを用いて医薬品ノンコンプライアンスの発言を分析するモデルを構築する。このモデルは転移学習を活用し、限られたデータ環境下でも高い性能を示す。さらに、薬物乱用とデータ不足という2つの課題に取り組む。1つ目の課題については、ソーシャルメディアにおける薬物乱用の実態を分析する新しい分類フレームワークを開発し、乱用の傾向や影響を可視化する。2つ目の課題に対しては、生成モデルを用いたデータ拡張により、学習データ不足がモデル訓練に与える影響を解決する方法を提案する。生成されたデータと実際の医療統計を比較し、補完的な学習データとしての有用性を検証する。本論文では、非構造化データである医療文書とソーシャルメディアからの情報抽出とその利用に焦点を当て、症状や副作用の可視化、ノンコンプライアンスの分析、データ不足の克服といった課題への取り組んだ。情報不足への対策や薬物乱用の可視化といった課題への対応を通じて、医療の質と患者安全の向上に寄与することを目指している。