

論文内容の要旨

博士論文題目

A Study on Automating Meta-Analysis Statistical Analysis by Employing Natural Language Processing Techniques (自然言語処理による自動メタアナリシスに関する研究)

氏 名 MUTINDA FAITH WAVINYA

(論文内容の要旨)

メタアナリシスとは、医学研究において、異なる臨床試験の結果を集約し、効果を評価するものである。メタアナリシスは、多い場合は何百もの研究論文を読み、データを抽出するため、膨大な時間と労力がかかる。さらに、研究論文の数は急速に増加しており、ほとんどのメタアナリシスは出版後すぐに時代遅れになってしまう。そこで、本研究では、論文からデータを自動抽出することで、メタアナリシスのプロセスを迅速化し、半自動的に更新を可能にすることを目指す。

研究はそのための新たなデータ構築と、データを用いた自動抽出部分、最後にメタアナリシス可視化部分となる。

まず、乳がんランダム化比較試験の PubMed 抄録約 1000 件に、臨床試験のコアとなる要素を注釈したコーパスを作成した。このコーパスは、臨床試験の中核的要素である PICO (Participants, Intervention, Control, and Outcomes) を注釈として含んでいる。

次に、このコーパスを用いて、BERT に基づく情報抽出モデル (NER モデル) を提案した。NER モデルはほとんどのエンティティで 0.80 を超える F1 スコアを達成するなど、比較的高い精度で PICO 要素を抽出可能なものである。

さらに、これを総合して、メタアナリシス統計解析を行うための可視化システムを構築した。

氏名	MUTINDA FAITH WAVINYA
----	-----------------------

(論文審査結果の要旨)

本研究は、医学論文のアブストラクトからデータを自動抽出することで、メタアナリシスのプロセスを半自動的に実現するシステムを構築した。

メタアナリシスの自動化のためには、構成要素である臨床試験の中核的要素であるPICO (Participants, Intervention, Control, and Outcomes)を機械が抽出し、解釈する必要がある。これを高い精度で抽出し、さらに、集約するための可視化システムを構築した。

研究が細分化、かつ加速しつつある医学分野におけるメタアナリシスは、多い場合には何百もの研究論文を読み、データを抽出するため、膨大な時間と労力がかかる。さらに、多くのメタアナリシスは出版後、数年も立たずして時代遅れになってしまう。このような状況の中、メタアナリシス実行コストを軽減する本研究の意義は極めて高い。よって、博士(工学)の学位論文としての価値があるものと認める。