Graduate School of Information Science   Doctoral Thesis Abstract

Lab name (Supervisor)	Social Computing Professor Aramaki Eiji		
Name (surname) (given name)	Mutinda Faith Wavinya	Date	(2021/11/22)
Title	A Study on Automating Meta-Analysis Statistical Analysis by Employing Natural Language Processing Techniques		

Abstract

Background and objective: Meta-analyses aggregate results of different clinical studies to assess the effectiveness of a treatment. Despite their importance, meta-analyses are time-consuming and labor-intensive as they involve reading hundreds of research articles and extracting data. The number of research articles is increasing rapidly and most meta-analyses are outdated shortly after publication as new evidence has not been included. Automatic extraction of data from research articles can expedite the meta-analysis process and allow for automatic updates when new results become available. In this study, we propose a system for automatically extracting data from research abstracts and performing statistical analysis.

Materials and methods: We created a corpus consisting of 1011 PubMed abstracts of breast cancer randomized controlled trials annotated with the core elements of clinical trials: Participants, Intervention, Control, and Outcomes (PICO). We proposed a BERT-based named entity recognition (NER) model to identify PICO information from research abstracts. After extracting the PICO information, we parse numeric outcomes to identify the number of patients having certain outcomes for statistical analysis.

Results: The NER model extracted PICO elements with relatively high accuracy, achieving F1-scores greater than 0.80 in most entities. We assessed the performance of the proposed system by reproducing the results of an existing meta-analysis. The data extraction step achieved high accuracy, however the statistical analysis step achieved low performance because abstracts sometimes lack all the required information.

Conclusion: We proposed a system for automatically extracting data from research abstracts and performing statistical analysis. We evaluated the performance of the system by reproducing an existing meta-analysis and the system achieved a relatively good performance, though more substantiation is required.