Information Extraction from Unstructured Text using Large Language Models for Natural Product-Drug Interactions

Sanya B. Taneja, Sonish Sivarajkumar, Yanshan Wang, Richard D. Boyce
University of Pittsburgh, Pittsburgh, PA, USA

NaPDI
Center of Excellence for Natural Product-Drug Interaction Research

INTRODUCTION

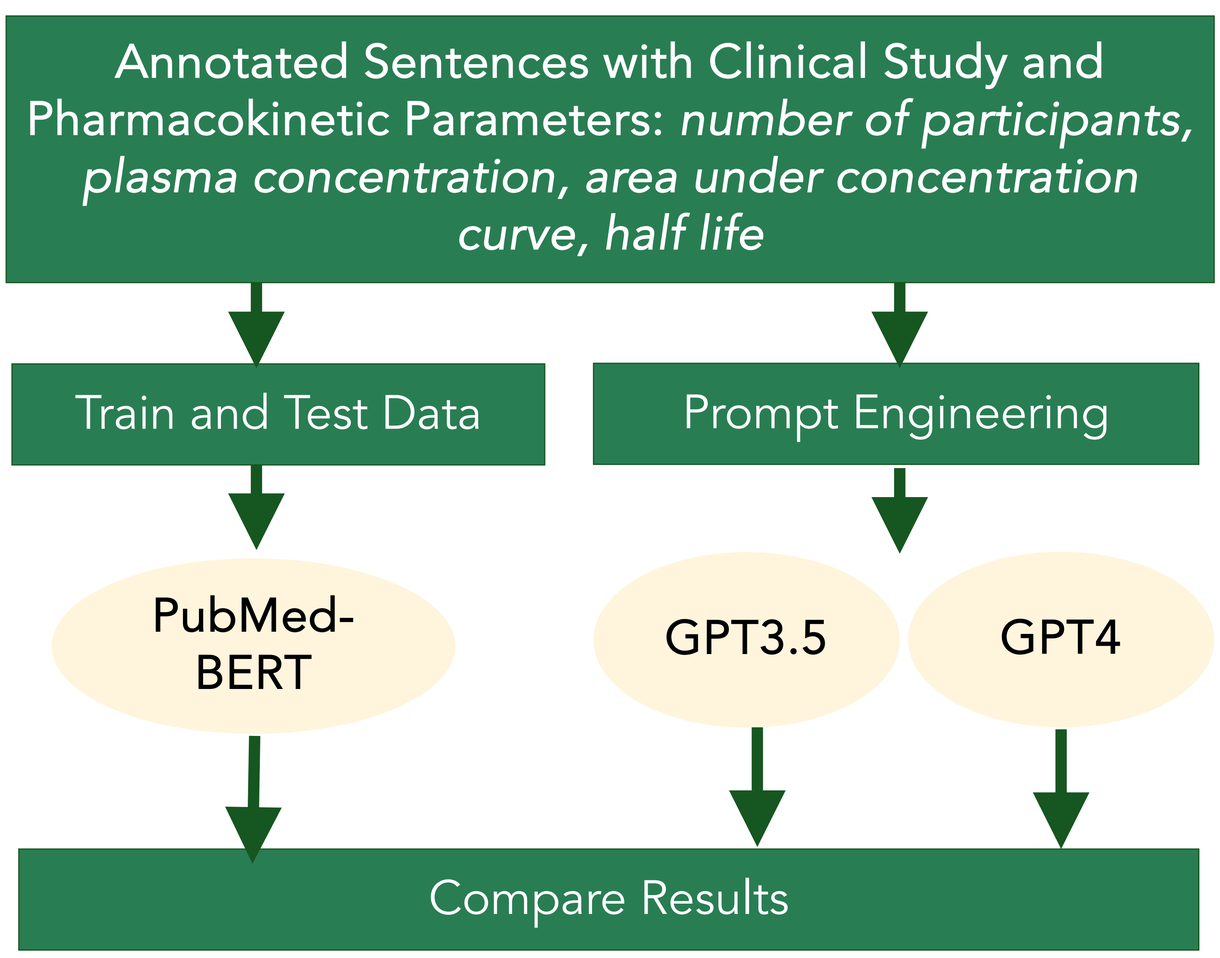
MOTIVATION

LLMs can be used for information extraction beyond NER and relation extraction, including context associated with events in literature-based discovery and aggregation of data from clinical studies.

OBJECTIVES

- Evaluate LLMs to extract important parameters from clinical studies from unstructured text and tables.
- Assess GPT models' effectiveness in annotating sentences from literature to reduce annotation burden.

METHODS



LLM: Large Language Models
GPT: Generative Pretrained Transformer
GPT3.5 and GPT4: GPT models developed by OpenAI

RESULTS

	No. of parameters	No. of sentences (%)
Number of participants	26	25 (47.2)
Plasma concentration	30	24 (45.3)
Area under conc. curve	26	23 (43.4)
Half life	3	3 (5.7)
Total	85	53 (100)

Table 1: Parameters and sentences in test data.

	PubMed-BERT	GPT3.5	GPT4
Accuracy (%)	92.5	54.72	77.36
F1-score	0.70	-	-

Table 2: Performance of large language models on test data.

Prompt:

```
The task is to extract pharmacokinetic study measurements from the given sentence.
For the sentence delimited by <>, do the following -
1. extract measurement values in the sentence in JSON format like
   [measurement_type: [(measurement value 1 with units, position of measurement
   value 1).
2. return only the JSON output without explanation.
3. If there are no measurement values in the sentence, return 'no measurements' as
   the JSON key.
measurement_type should be values from the set (number of participants or
subjects, area under the plasma concentration curve (AUC), maximum plasma
concentration (Cmax), increase/decrease in plasma concentration/levels). 'position
of measurement value' are the start and end character indices of the measurement
value in the sentence and characters are counted sequentially from the beginning
of the sentence including punctuation and spaces as characters.
```

Common Errors with GPT models:

- Hallucination of parameter types not in the prompt – more frequent in GPT3.5
- Character spans not provided accurately.

CONCLUSION

- Although GPT-4 can retrieve relevant information from limited context, manual review will be required when using it for data annotation.
- Future experiments will explore open-source models such as Llama2, information extraction with text and tables, extract additional parameters (e.g., dosage), and aim to enhance PubMedBERT's performance with annotated data and prompting strategies from LLMs.

Gu Y, Tinn R, Cheng H, Lucas M, Usuyama N, Liu X, Naumann T, Gao J, Poon H. Domain-specific language model pretraining for biomedical natural language processing. ACM Transactions on Computing for Healthcare (HEALTH). 2021 Oct 15;3(1):1-23.
Sivarajkumar S, Kelley M, Samolyk-Mazzanti A, Visweswaran S, Wang Y. An empirical evaluation of prompting strategies for large language models in zero-shot clinical natural language processing. arXiv preprint arXiv:2309.08008. 2023 Sep 14.
https://platform.openai.com/docs/models

This work was funded by the National Institutes of Health National Center for Complementary and Integrative Health [Grant U54 AT008909].