How well ChatGPT understand Malaysian English? An Evaluation on Named Entity Recognition and Relation Extraction

Mohan Raj Chanthran*, Lay-Ki Soon, Huey Fang Ong, Bhawani Selvaretnam

*mohan.chanthran@monash.edu

MONASH University

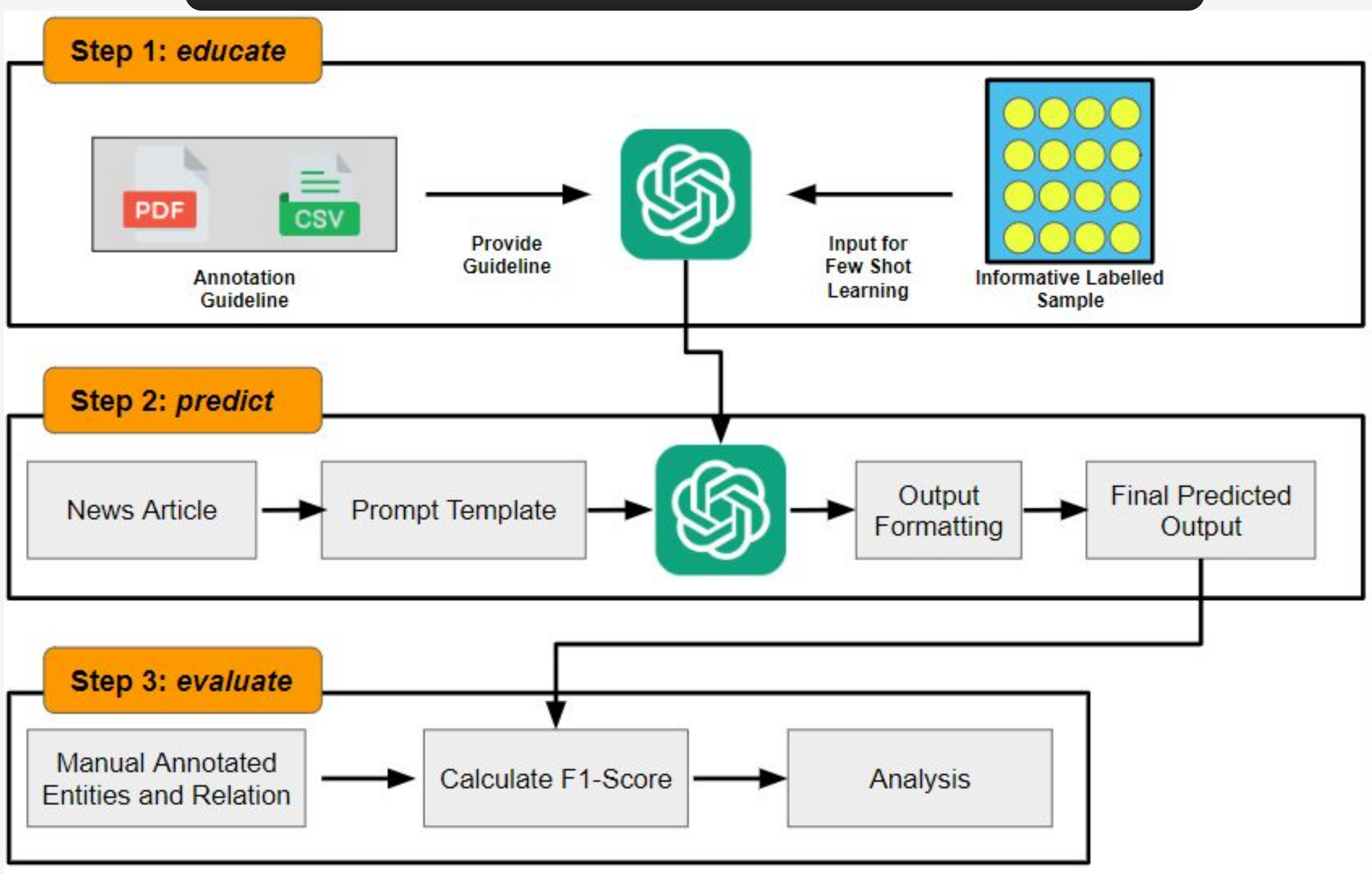
What is Malaysian English

Malaysian English (ME) has evolved into a unique form of English incorporating local words from languages like Bahasa Malaysia, Chinese and Tamil.

Introduction

1. ChatGPT has demonstrated strong performance across various NLP downstream tasks in Standard English.
2. However, this has raised a question ***How effective is ChatGPT capable of extracting entities and relations from Malaysian English News?***
3. This question has been raised as Malaysian English exhibits morphosyntactic adaptations like usage of **loan words, compound blend and derived words**.

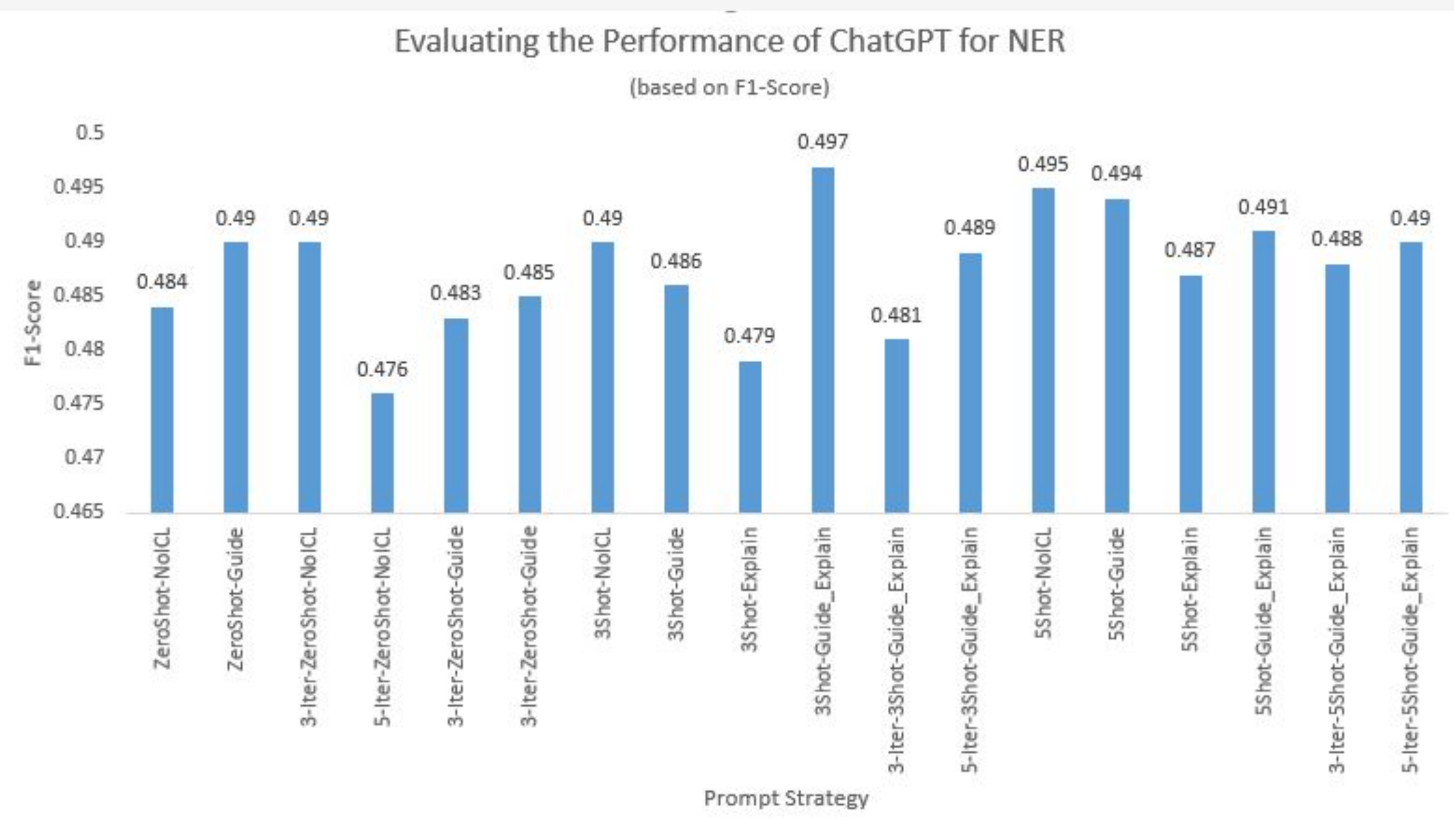
educate-predict-evaluate



1. ***educate***: Enhancing ChatGPT by enabling **In-Context Learning** (ICL) with annotation guideline.
2. ***predict***: Predicting entities and relations using ChatGPT with different prompting techniques like, **Zero Shot Prompting, Few-Shot Prompting and Few-Shot with Explanation Prompting**.
3. ***evaluate***: ChatGPT's NER and RE **performance** were rigorously evaluated via **F1-Score** calculations, benchmarked **against human annotations**.

There are 18 different prompt settings has been used in this evaluation. Those 18 different prompt settings are based on different ICL, and prompting techniques.

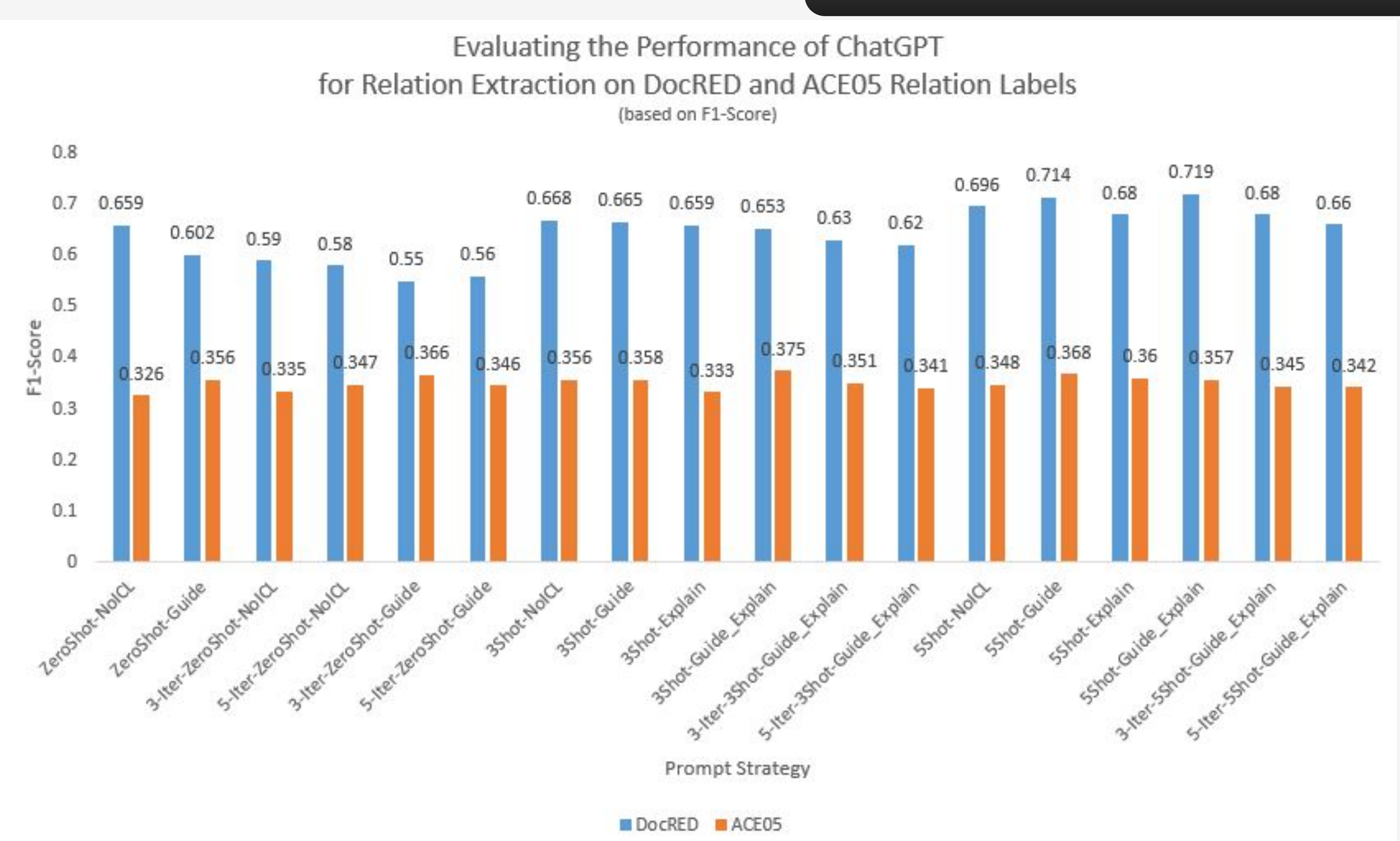
Evaluating ChatGPT for Named Entity Recognition (NER)



1. ***How well did ChatGPT perform in extracting entities from Malaysian English? Does it perform better?***
   a. Although various prompting techniques has been evaluated, the overall difference of F1-Score recorded is 0.488 ± 0.01. The highest F1-Score is 0.497.
   b. During MEN-Dataset annotation, the Inter-Annotator Agreement score is 0.81, while ChatGPT highest F1-Score of 0.497, revealing performance limitations.
2. ***What are the limitations of ChatGPT in extracting entities? Were there specific types of entity labels that ChatGPT consistently struggled to extract or misidentify?***
   a. For entity label PERSON, we noticed most errors due to exists of Compound Blend .
   b. ChatGPT not extracting abbreviations of ORGANIZATION.
   c. For NORP, we noticed most of the errors as they are Derived Words.

**Scan Result and Analysis QR for more details

Evaluating ChatGPT for Relation Extraction (RE)



1. ***How accurate was ChatGPT in extracting relations between entities, and were there any notable errors or challenges?***
   a. Average F1-Score for relation adapted from DocRED and ACE05 are 0.64 and 0.35 respectively. This gap is due to complexity in understanding relation labels.
   b. In-Context Learning improved the performance of ChatGPT in identifying the relations.
   c. 5 Shot Learning slightly improved the performance of ChatGPT, compared to 3 Shot Learning of various prompting techniques.
   d. Although no morphosyntactical adaptation is required for predicting relations, the performance of ChatGPT in relation prediction is influenced by its understanding of the context within the news article

**Scan Result and Analysis QR for more details

Conclusion

1. Experiment results show morphosyntactic adaptation significantly influenced ChatGPT's entity extraction in Malaysian English news articles.
2. As future work, we will expand the experiment to various LLM and various downstream tasks.

arXiv Paper

Connect in LinkedIn

Result and Analysis