NGNE - Final Report

Bao Honghui

School of Computing, National University of Singapore

April 21, 2023

Abstract

Use **Prompt Engineering** to design the best prompts to drive **GPT-3** to produce the best results

Introduction - Brief Background

- **Document-Level Relation Extraction**
  Focuses on identifying relations between entities within multiple sentences.
- **GPT-3**
  A very powerful large-scale pre-trained language model.
- **Prompt Engineering**
  Designing and optimizing prompts for NLP models to have the desired response.

Introduction - Motivation

- **Room for Exploration in the Task**
  Document-Level Relation Extraction.
- **GPT Series**
  Powerful!
- **Prompt + GPT**
  Best Prompt leads to Best Performance.

Detailed Background - Document-level Relation Extraction

DocRED (Document-level Relation Extraction Dataset):

- Many relational facts are expressed across multiple sentences.
- There are few general dataset for this task.

Re-DocRED (Revisiting Document-level Relation Extraction Dataset):

- Incomplete Annotations: False Negatives be prevalent.

Detailed Background - LLM

- Strong Generalization Ability and Robustness.
- Handle Complex Language Structure and Semantic Information.
- Zero-shot

Detailed Background - Prompt Engineering

Def. Crafting effective and precise prompts to train natural language processing models.

Simple Examples of Prompt Learning:

- Sentiment Analysis - "I like this movie."
- Prompt Engineering - "I like this movie, overall, it's a __ movie."
- Answer Mapping - Output

Related Work - Two Types of Prompt Engineering

Manual Template Engineering:

- Cloze Prompt - LAMA Dataset[3].
- Prefix Prompt - Create hand-crafted prompts[1].

Automated Template Engineering:

- Discrete Prompt - "Gradient-based" [4]
- Continuous Prompt - "P-tuning" [2]

Prompts Engineering - Recap of Dataset

The Data Structure:

1. Data

Data Format:

- **title** : The title of article
- **sents:** List of words in each sentence in the article.
- **vertexSet:**
  - "name": An entity
  - "sent_id": The id of sentence which contains the entity. (For one sentence)
  - "pos": The start and end position of the entity in the sentence.
  - "type": The type of entity.
- **labels**
  - 'h': The index position of the first entity in vertexSet
  - 't': The index position of the other entity in vertexSet
  - 'r': The index of **relation**
  - 'evidence': The id of supporting evidence sentences.

Figure: Structure of Re-DocRED

Prompts Engineering - Basic Prompt

Basic Prompt:

```
Instruction: the paragraph is from an article of Wikipedia.

 Specify the relations that exist between the entity "**entity1**" and entity "**entity2**" based on the paragraph.

**relation**

"""
**paragraph**
"""
```

Figure: Basic Prompt

Prompts Engineering - Enhanced Prompt 1: Attention

Attention: You are restricted to choose from the following relations and if there are matching relationships below, please output the relations directly, otherwise output "NONE":

Prompts Engineering - Enhanced Prompt 2: Grouping

We have 96 relations in the dataset.

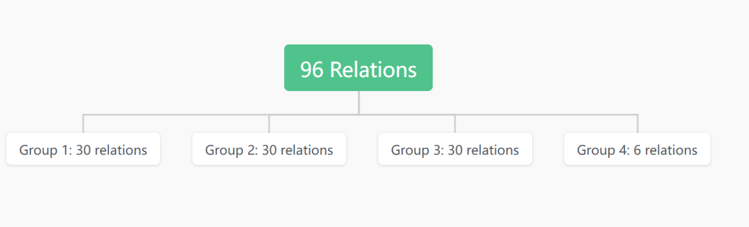

Figure: Grouping

Prompts Engineering - Enhanced Prompt 3: Shuffle

P569	date of birth	date on which the subject was born
P570	date of death	date on which the subject died
P571	inception	date or point in time when the organization/subject was founded/created
P576	dissolved, abolished or demolished	date or point in time on which an organisation was dissolved/disappeared or a building demolished; see also discontinued date (P2669)
P577	publication date	date or point in time a work is first published or released
P580	start time	indicates the time an item begins to exist or a statement starts being valid
P582	end time	indicates the time an item ceases to exist or a statement stops being valid
P585	point in time	time and date something took place, existed or a statement was true

Figure: Relation Cluster

Prompts Engineering - Enhanced Prompt 4: Order

The order of paragraph:

```
Instruction: the paragraph is from an article of Wikipedia.

"""
**paragraph**
"""

 Specify the relations that exist between the entity "**entity1**" and entity "**entity2**" based on the paragraph.
Attention: You are restricted to choose from the following relationships and if there are matching relationships below,
please output the relations directly, otherwise output "NONE":

**relation**
```

Figure: Order of paragraph

Prompts Engineering - Enhanced Prompt 4: Order

The order of entity:

```
Instruction: the paragraph is from an article of Wikipedia.

"""
**paragraph**
"""
For entity "**entity1**", what relationship is entity "**entity2**" to it?
Attention: You are restricted to choose from the following relationships and if there are matching relationships below,
please output the relations directly, otherwise output "NONE":

**relation**
```

Figure: Order of Entity

('Rihanna', 'Loud', ['notable work'])
('Loud', 'Rihanna', ['performer'])

Prompts Engineering - Other Attempts

- Name Entity Types
- Relation Explanations
- ChatGPT
- ...

Experiment - Setting

Setting

- Jupyter Notebook
- Python 3.7.9
- GPT-3 API

GPT-3 API

Parameter	**Value**
engine	"text-davinci-003"
temperature	0.1
max length	256
top_p	1
frequency_ penalty	0

Experiment - Process

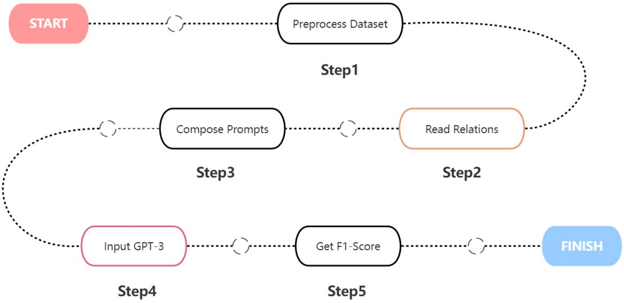

Figure: Experimental Process

Experiment - Dataset

- **Motivation**: Released Time.
- **How to Build**:
  - Define the relation.
  - Identify the Documents.
  - Annotate the Data.
- **Analysis**: 5 Documents

Experiment - Metric

F1-Score:

$$F1 = 2 * (Precision * Recall)/(Precision + Recall)$$

Precision and Recall:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Experiment - ATLOP

ATLOP[5] is one of the SOTA models in the document-level relation extraction task.

Experiment - Comparison

Model	Re-DocRED			Self-Made Dataset		
	Prec	Rec	F1	Prec	Rec	F1
ATLOP	0.93	0.67	0.78	0.5	0.32	0.39
Basic Prompt	0.09	0.74	0.17	0.06	0.24	0.10
Enhanced Prompt 1 (attention)	0.31	0.26	0.28	0.10	0.02	0.03
Enhanced Prompt 2 (group)	0.26	0.36	0.31	0.10	0.12	0.11
Enhanced Prompt 3 (order-paragraph)	0.26	0.56	0.36	0.25	0.39	0.31
Enhanced Prompt 4 (shuffle)	0.35	0.79	0.49	0.44	0.78	0.56
Enhanced Prompt 5 (order-entity)	0.36	0.82	0.50	0.41	0.80	0.55

Experimental results

Experiment - Comparison

Comparison between Basic Prompt and Enhanced Prompts:

- Enhanced Prompt 1: Attention
- Enhanced Prompt 2: Grouping
- Enhanced Prompt 3: Shuffle
- Enhanced Prompt 4: Order

Comparison between Best-Performing Prompt and ATLOP

Experiment - Comparison Example

Paragraph:
"The Loud Tour was the fourth overall and third world concert tour by Barbadian recording artist Rihanna. Performing in over twenty countries in the Americas and Europe , the tour was launched in support of Rihanna's fifth studio album Loud (2010) "
Specify the relation between entity "Loud" and "2010":

- **"publication date"**
- **"inception"**

Experiment - Discussion

1. **What's the most important component to the prompt?**

Fromt the results:

- Attention
- Shuffle

2. **Precision's Problem**

Conclusion

Remaining Problem:

- Zero-shot vs Few-shots
- Balance between Precision and Recall

Future Work:

- Relations' Information.
- Complex GPT Model Exploration.

Thank you!

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei.
Language models are few-shot learners, 2020.

Xiao Liu, Yanan Zheng, Zhengxiao Du, Ming Ding, Yujie Qian, Zhilin Yang, and Jie Tang.
Gpt understands, too, 2021.

Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander Miller.
Language models as knowledge bases?
In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International*

Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 2463–2473, Hong Kong, China, November 2019. Association for Computational Linguistics.

Eric Wallace, Shi Feng, Nikhil Kandpal, Matt Gardner, and Sameer Singh.
Universal adversarial triggers for attacking and analyzing NLP.
In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2153–2162, Hong Kong, China, November 2019. Association for Computational Linguistics.

Wenxuan Zhou, Kevin Huang, Tengyu Ma, and Jing Huang.
Document-level relation extraction with adaptive thresholding and localized context pooling, 2020.