Language Model Prompting

Wei-Rui Chen

Outline

1. What is prompt?
   a. Formulation
   b. Examples
2. Engineering toward better performance
   a. Choosing suitable pre-trained language model (PLM)
   b. Prompt engineering
   c. Answer engineering
   d. Multi-prompt
3. Recent Work (ACL 2022)
   a. Jin et al 2022
   b. Lu et al 2022

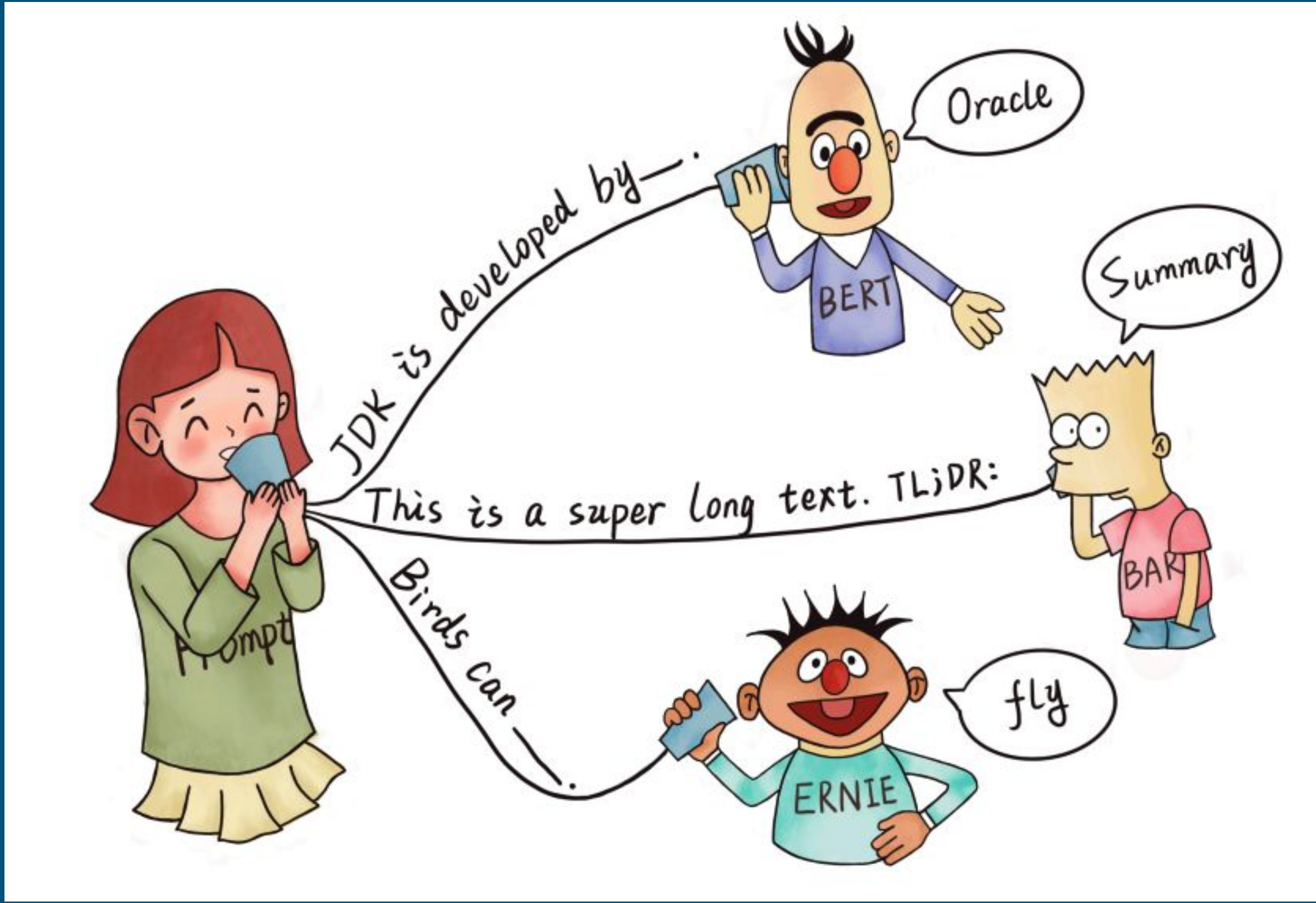
1. What is Prompt

Ways to do NLP tasks

Non-neural network: feature engineering

Neural network

- architecture engineering - CNN, RNN, Transformer
- Pretrained -> Fine-tuned - BERT, GPT
- Pretrained -> Prompt



Source: Liu et al. 2021

Formulation

Name	Notation	Example	Description
Input	$\boldsymbol{x}$	I love this movie.	One or multiple texts
Output	$\boldsymbol{y}$	++ (very positive)	Output label or text
Prompting Function	$f_{\text{prompt}}(\boldsymbol{x})$	`[X]` Overall, it was a `[Z]` movie.	A function that converts the input into a specific form by inserting the input $\boldsymbol{x}$ and adding a slot `[Z]` where answer $\boldsymbol{z}$ may be filled later.
Prompt	$\boldsymbol{x}'$	I love this movie. Overall, it was a `[Z]` movie.	A text where `[X]` is instantiated by input $\boldsymbol{x}$ but answer slot `[Z]` is not.
Filled Prompt	$f_{\text{fill}}(\boldsymbol{x}', \boldsymbol{z})$	I love this movie. Overall, it was a bad movie.	A prompt where slot `[Z]` is filled with any answer.
Answered Prompt	$f_{\text{fill}}(\boldsymbol{x}', \boldsymbol{z}^*)$	I love this movie. Overall, it was a good movie.	A prompt where slot `[Z]` is filled with a true answer.
Answer	$\boldsymbol{z}$	"good", "fantastic", "boring"	A token, phrase, or sentence that fills `[Z]`

Table 2: Terminology and notation of prompting methods. $\boldsymbol{z}^*$ represents answers that correspond to true output $\boldsymbol{y}^*$.

Source: Liu et al. 2021

Formulation

$$\mathcal{Z} = \{\text{“excellent”}, \text{“good”}, \text{“OK”}, \text{“bad”}, \text{“horrible”}\}$$

$$\hat{\boldsymbol{z}} = \operatorname*{search}_{\boldsymbol{z} \in \mathcal{Z}} P(f_{\text{fill}}(\boldsymbol{x}', \boldsymbol{z}); \theta)$$

Source: Liu et al. 2021

Examples

Type	Task	Input ([X])	Template	Answer ([Z])
Text CLS	Sentiment	I love this movie.	[X] The movie is [Z].	great fantastic ...
	Topics	He prompted the LM.	[X] The text is about [Z].	sports science ...
	Intention	What is taxi fare to Denver?	[X] The question is about [Z].	quantity city ...
Text-span CLS	Aspect Sentiment	Poor service but good food.	[X] What about service? [Z].	Bad Terrible ...
Text-pair CLS	NLI	[X1]: An old man with ... [X2]: A man walks ...	[X1]? [Z], [X2]	Yes No ...
Tagging	NER	[X1]: Mike went to Paris. [X2]: Paris	[X1][X2] is a [Z] entity.	organization location ...
Text Generation	Summarization	Las Vegas police ...	[X] TL;DR: [Z]	The victim ... A woman ... ...
	Translation	Je vous aime.	French: [X] English: [Z]	I love you. I fancy you. ...

Table 3: Examples of *input*, *template*, and *answer* for different tasks. In the **Type** column, "CLS" is an abbreviation for "classification". In the **Task** column, "NLI" and "NER" are abbreviations for "natural language inference" (Bowman et al., 2015) and "named entity recognition" (Tjong Kim Sang and De Meulder, 2003) respectively.

Source: Liu et al. 2021

2. Engineering toward better performance

Choosing suitable PLM

Autoregressive LM -> NLG

$$P(\boldsymbol{x}) = P(x_1) \times \cdots P(x_n | x_1 \cdots x_{n-1})$$

Masked LM -> NLU

$$P(x_i | x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

Source: Liu et al. 2021

Prompt Engineering - Prompt Shape

Prefix

… [X] … [Z]

[X] The movie is [Z]

Cloze

…[X] … [Z] …

[X] is a [Z] entity

Prompt Engineering - Automatic/Manual Template Engineering

Manual Template Engineering

Pros: Simplicity

Cons: Human effort, can be suboptimal

Automatic Template Engineering

Pros: Less human effort

Cons: Relatively complicated

Prompt Engineering - Automatic Prompt Engineering

Prompt Mining

Let x (e.g. I like the movie) be input and y be target (e.g. great). Scrape large corpus and find “middle words” between these two terms. Design a template based on the observations.

Prompt Paraphrasing

Get paraphrases of a existing template, and select the template with highest training accuracy.

Answer Engineering - Answer Shape

Tokens

e.g. Text classification

Multi-token span

e.g. QA

Sentence

e.g. Summarization

Answer Engineering - Automatic/Manual

Manual

Unconstrained / constrained answer space (V or subset of V)

Automatic

Answer paraphrasing

$$P(\boldsymbol{y}|\boldsymbol{x}) = \sum_{\boldsymbol{z} \in \mathrm{para}(\boldsymbol{z}')} P(\boldsymbol{z}|\boldsymbol{x})$$

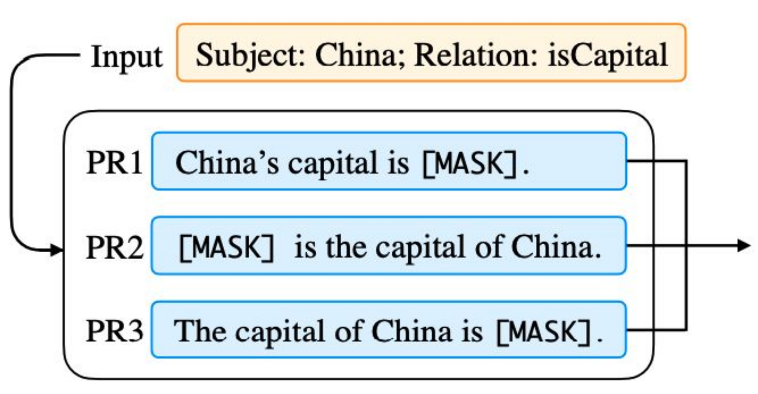
Source: Liu et al. 2021

Multi-Prompt - Prompt Ensembling

Prompt Ensembling: Make prediction with multiple unanswered prompts

Pros:

- leverage the complementary advantages of different prompts
- Reduce the effort to search for the best
- stabilize performance on downstream tasks



Source: Liu et al. 202

Multi-Prompt - Prompt Ensembling

How to take average among multiple unanswered prompts

Uniform averaging

$$P(\boldsymbol{z}|\boldsymbol{x}) := \frac{1}{K}\sum_{i}^{K} P(\boldsymbol{z}|f_{\text{prompt},i}(\boldsymbol{x}))$$

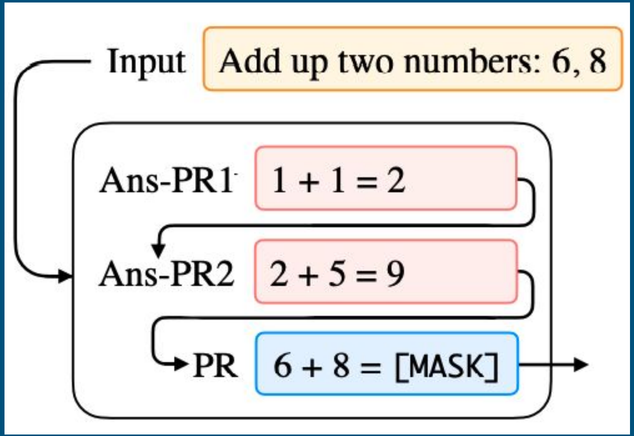
Weighted averaging

Weighted with each prompts' performance

Multi-Prompt - Prompt Augmentation

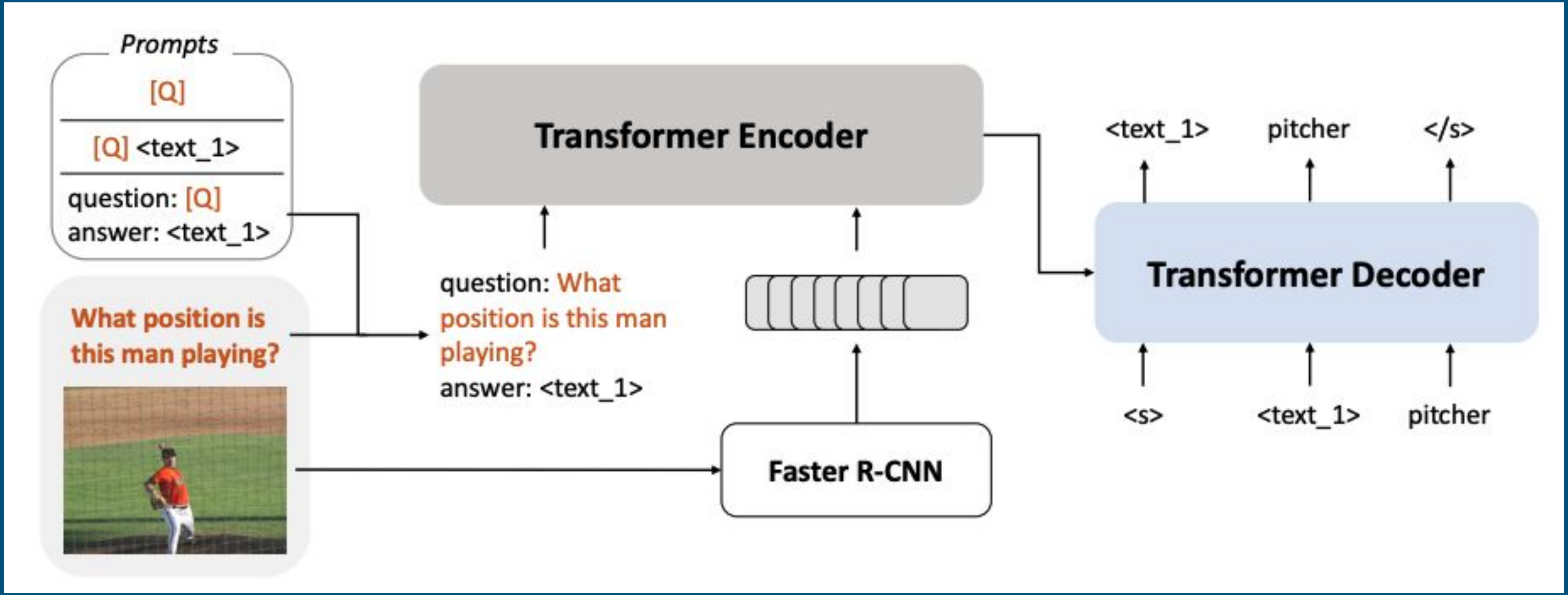
Show PLM a few answered prompt before asking it to predict

Great Britain's capital is London . Japan's capital is Tokyo . China's capital is [Z]



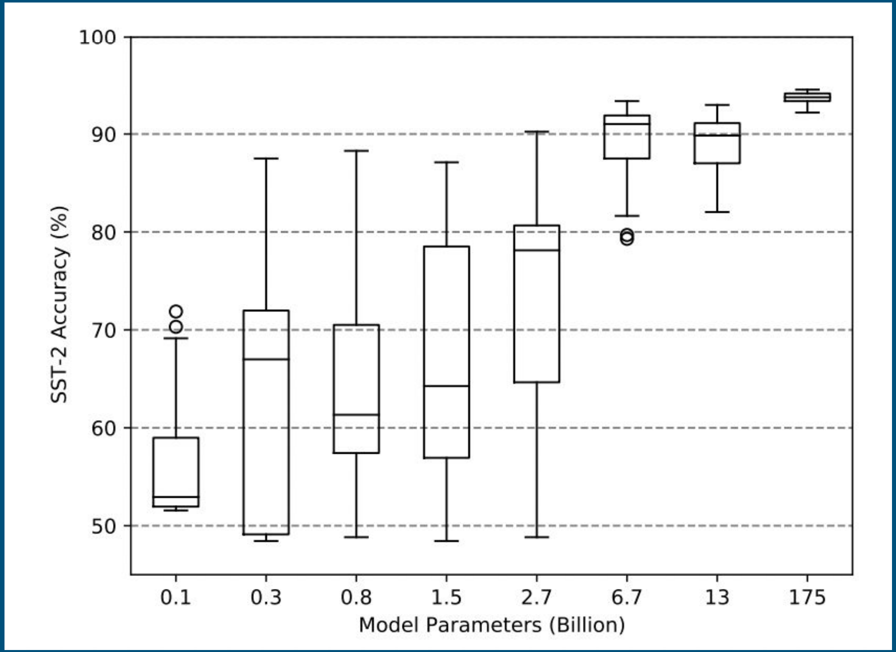
3. Recent Work

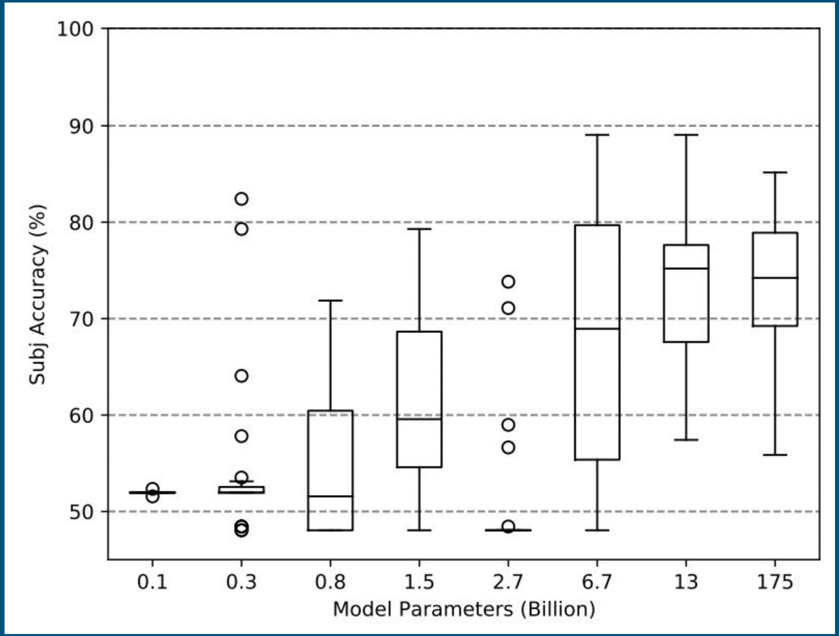
Jin et al 2022



Source: Jin et al. 2022

Lu et al. 2022 - Sample Order Matters





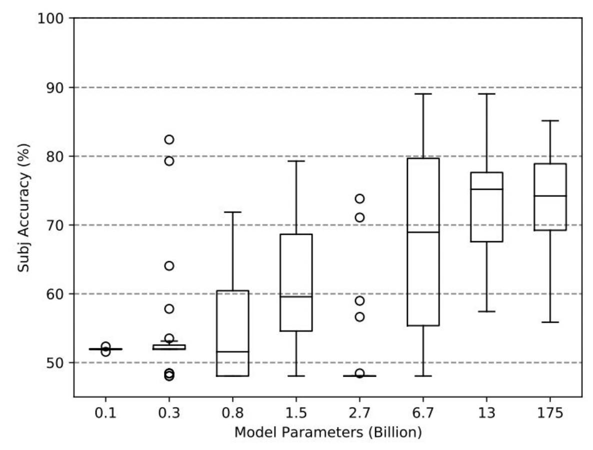
Source: Lu et al. 2022

Lu et al. 2022 - Sample Order Matters

Increasing model size does not guarantee low variance

Performant permutation are not transferable across models

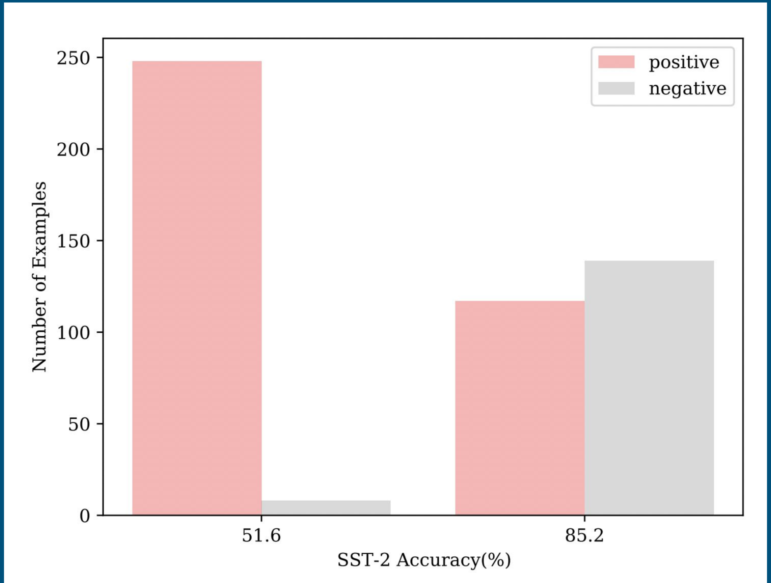
88.7% acc in GPT2-XL -> 51.6% acc in GPT2-Large



Source: Lu et al. 2022

Lu et al. 2022 - Methodology

Error analysis: bad prompt <-> unbalanced predicted label distribution

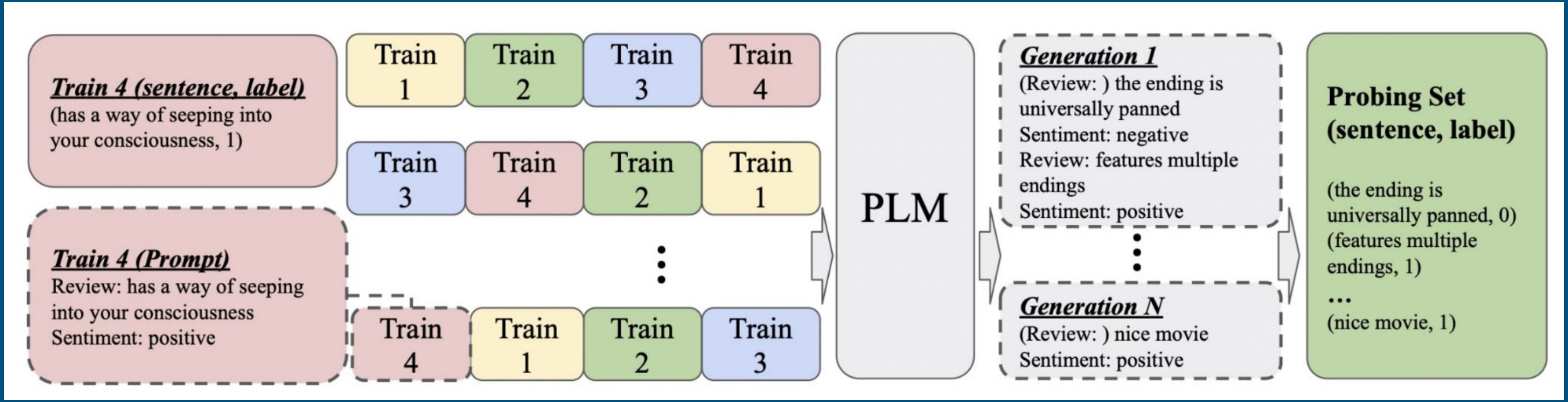


Source: Lu et al. 2022

Lu et al. 2022 - Methodology

Step 1: Find the best among n! probing sets with highest entropy

Step 2: inverse function to get the best sample ordering



Source: Lu et al. 2022

Q & A