

Carnegie Mellon University Language Technologies Institute



The Fourth Paradigm of Modern Natural Language Processing Techniques

Pengfei Liu



Outline

- What is the "Prompt"?
- What is the general workflow of prompt-based methods?
- What are the design considerations for prompt-based methods?
- What (unique) advantages could prompt learning bring to us?
- How does prompt-based research progress currently?

What is the "Prompt"?

Prompt meaning prompt <

Words form:

prompted promptest prompting prompts

See word origin >

The definition of a prompt is a cue given to someone to help him verb remember what to say, or is something that causes another event or action to occur.

An example of prompt is when you whisper a line to an actor who forgot what to say next.

An example of prompt is an event that starts an argument.

Google	Q what are the most bea	X 🎙 🔍
	Q what are the most beautiful names	
Q All 🔳 Boo	Q what are the most beautiful places in the world	
About 7,420,000,	Q what are the most beautiful zodiac signs	Prompts
	Q what are the most beautiful flowers	

What is the "prompt" in the context of NLP research?

An Intuitive Definition

Prompt is a cue given to the pre-trained language model to allow it better understand human's questions



More Technical Definition

Prompt is the technique of making better use of the knowledge from the pre-trained model by adding additional texts to the input.

More Technical Definition

Purpose

Prompt is the technique of making better use of the knowledge from the pre-trained model by adding additional texts to the input.



What is the general workflow of prompt-based methods?

Workflow for Prompting Methods

- Prompt Construction
- Answer Construction
- Answer Prediction
- Answer-Label Mapping

Prompting for Sentiment Classification

Task Description:

□ **Input**: sentence x;

□ **Output**: emotional polarity of it

(i.e.,☺ v.s ອ).

Input: x = I love this movie.

Step 1: Prompt Construction

 Transform x into prompt x' through following two steps:

□ Defining a **template** with two **slots**:

[x] and [z];



Step 1: Prompt Construction

 Transform x into prompt x' through following two steps:



□ Defining a **template** with two **slots**:

[x] and [z];

Template: [x] Overall, it was a

Input: x = I love this movie.

[z] movie.

Require human effort

Step 1: Prompt Construction

 Transform x into prompt x' through following two steps:



Defining a template with two slots:
[x] and [z];

 \Box Instantiate slot [x] with input text.



Step 2: Answer Construction

Build a mapping function between answers and class labels.





Step 3: Answer Predicting

Given a prompt, predict the answer [z].



□ Choose a suitable pretrained

language model;



Step 3: Answer Predicting

Given a prompt, predict the answer [z]



 $\hfill\square$ Choose a suitable pretrained

language model;



□ Fill in [z] as "fantastic"



Step 4: Answer Mapping

 Mapping: Given an answer, map it into a class label.

□ fantastic => ⓒ



Terminology	Notation	Example	
Input	X	I love this movie	
Output (label)	У		
Template	-	[x] Overall, it was a [z] movie	
Prompt	х'	I love this movie. Overall, it was a [z] movie	
Answer	Z	fantastic, boring	

Rethinking Human Efforts in Prompt-based Methods



Rethinking Human Efforts in Prompt-based Methods



What are the design considerations for promptbased methods?

Design Considerations for Prompt-based Methods

- Prompt Template Engineering
- Answer Engineering
- Pre-trained Model Choice
- Expanding the Paradigm
- Prompt-based Training Strategies

Design Considerations for Prompt-based Methods

Prompt Template Engineering

- Answer Engineering
- Pre-trained Model Choice
- Expanding the Paradigm
- Prompt-based Training Strategies

Prompt Template Engineering

Research Question:

□ how to define appropriate prompt templates

It was a [z] movie



Design Decision of Prompt Templates



One Example



One Example



One Example



Design Considerations for Prompt-based Methods

Prompt Template Engineering

Answer Engineering

- Pre-trained Model Choice
- Expanding the Paradigm
- Prompt-based Training Strategies

Answer Engineering

Research Question:

□ Given a task (or a prompt), how to define a suitable mapping function between label space and answer space?

{fantastic:⁽ⁱ⁾, boring:⁽ⁱ⁾}



Answer Engineering

Research Question:

□ Given a task (or a prompt), how to define a suitable mapping function between

label space and answer space?



Design Decision of Prompt Answer Engineering



Task	Template	Answer		
Sentiment Classification	[x] the movie is [z]	great fantastic boring	Token	
			bounded	
			Manual	Search


Design Considerations for Prompt-based Methods

- Prompt Template Engineering
- Answer Engineering
- Pre-trained Model Choice
- Expanding the Paradigm
- Prompt-based Training Strategies

Pre-trained Model Choice

Research Question:

□ Given a task (or a prompt), which pre-trained language model would be the

most appropriate one?



The story describes, in summary [z]



Design Decision of Pre-trained Models



Design Decision of Pre-trained Models



Left-to-right Language Model

Characteristics

□ First proposed by Markov (1913)

□ Count-based-> Neural network-based

□ Specifically suitable to highly larger-scale LMs

Example

□ GPT-1,GPT-2,GPT-3

Roles in Prompting Methods

□ The earliest architecture chosen for prompting

□ Usually equipped with prefix prompt and the parameters of PLMs are fixed



Masked Language Model

Characteristics

□ An extension of left-to-right architecture

□ Unidirection -> bidirection prediction

□ Suitable for NLU tasks

Example

□ BERT, ERNIE

Roles in Prompting Methods

□ Usually combined with cloze prompt

 \Box Suitable for NLU tasks



Encoder-Decoder

Characteristics

 \square A denoised auto-encoder

□ Use two Transformers and two different mask

mechanisms to handle text X and Y separately

Examples

 \Box BART, T5

Roles in Prompting methods

□ Text generation tasks or some tasks that can be

formulated into a text generation problem



Design Considerations for Prompt-based Methods

- Prompt Template Engineering
- Answer Engineering
- Pre-trained Model Choice
- Expanding the Paradigm
- Prompt-based Training Strategies

Expanding the Paradigm

Research Questions

□ How to extend the current prompting framework to support more NLP tasks?

Design Decision of Multiple Prompt Learning



Definition

using multiple unanswered prompts for an input at inference time to make predictions

Advantages

□ Utilize complementary advantages

□ Alleviate the cost of prompt engineering

□ Stabilize performance on downstream tasks



Prompt Augmentation

Definition

 Help the model answer the prompt with additional answered prompts

Advantage

 $\hfill\square$ make use of the small amount of information

that has been annotated

Core step

□ Selection of answered prompts

□ Ordering of answered prompts



Prompt Composition

Definition

 Prompts for a composable task can be designed with multiple sub-prompts, which can then be combined to complete the task

Advantage

□ It provides a method of prompt learning

for complex tasks



Prompt Decomposition

Definition

For tasks where multiple predictions
should be performed for one sample,
handle it individually

Advantages

 Break-down a complicated task into multiple separate ones



Definition

 When prompting method is applied to multiple tasks, domains or languages, prompts can be shared cross different tasks.

Advantage

□ Task- or language invariant information can be captured through prompting.



Design Considerations for Prompt-based Methods

- Prompt Template Engineering
- Answer Engineering
- Pre-trained Model Choice
- Expanding the Paradigm
- Prompt-based Training Strategies

Prompt-based Training Strategies

Data Perspective

□ Zero-shot: without any explicit training of the LM for

the down-stream task

□ Few-shot: few training (e.g., 100) samples of

downstream tasks

□ Full-data: lots of training samples (e.g., 10K) of

downstream tasks

Parameter Perspective





Promptless Fine-tuning

Example: BERT for text classification



Fixed-prompt Tuning

Example: BERT + Discrete Prompt for text classification



Fixed-prompt Tuning

Example: BERT + Transferred Continuous Prompt for text classification



Prompt+LM Fine-tuning

Example: BERT + Continuous Prompt for text classification



Adapter Tuning

Example: BERT + Adapter for text classification



Tuning-free Prompting

Example: GPT3 + Discrete Prompts for Machine Translation



Tuning-free Prompting

Example: GPT3 + Continuous Prompts for Machine Translation



Fixed-LM Prompt Tuning

Example: BART + Continuous Prompts for Machine Translation

Promptless Fine-tuning Fixed-prompt Tuning Prompt+LM Fine-tuning Adapter Tuning Tuning-free Prompting Fixed-LM Prompt Tuning

If you have a highly large left-toright pre-trained language model (e.g., GPT3)

If you have few training samples?

If you have lots of training samples?

What (unique) advantages could prompt learning bring to us?

Four levels of vision



Level-1: Make All NLP Tasks as a Language Modeling Problem

Level-1: Make All NLP Tasks as a Language Modeling Problem

Pretrained Language models can be better utilized



68

Level-1: Make All NLP Tasks as a Language Modeling Problem

- Pretrained Language models can be fully utilized
- (Almost) all NLP tasks can be handled zeroshotly



Sentiment Classification

Level-1: Make All NLP Tasks as a Language Modeling Problem

- Pretrained Language models can be fully utilized
- (Almost) all NLP tasks can be handled zeroshotly



Named Entity Recognition

70

Level-1: Make All NLP Tasks as a Language Modeling Problem

- Pretrained Language models can be fully utilized
- (Almost) all NLP tasks can be handled zeroshotly



Text Summarization

Level-1: Make All NLP Tasks as a Language Modeling Problem

- Pretrained Language models can be fully utilized
- (Almost) all NLP tasks can be handled zeroshotly
- Better few-shot performance

Level-1: (Almost) All NLP Tasks as Language Modeling

- Pretrained Language models can be fully utilized
- (Almost) all NLP tasks can be handled zero shotly
- Better few-shot performance
- Make different tasks methodologicallyconnected available

Extractive [SQuAD]

Question: At what speed did the turbine operate? Context: (Nikola_Tesla) On his 50th birthday in 1906, Tesla demonstrated his 200 horsepower (150 kilowatts) 16,000 rpm bladeless turbine. ... Gold answer: 16,000 rpm

Abstractive [NarrativeQA]

Question: What does a drink from narcissus's spring cause the drinker to do? Context: Mercury has awakened Echo, who weeps for Narcissus, and states that a drink from Narcissus's spring causes the drinkers to "Grow dotingly enamored of themselves." ... Gold answer: fall in love with themselves

Multiple-Choice [ARC-challenge]

Question: What does photosynthesis produce that helps plants grow? Candidate Answers: (A) water (B) oxygen (C) protein (D) sugar Gold answer: sugar

Yes/No [BoolQ]

Question: Was America the first country to have a president? Context: (President) The first usage of the word president to denote the highest official in a government was during the Commonwealth of England ... Gold answer: no

Unified QA (Daniel et al 2020) Different QA tasks are trained using one model
Level-2: Reverse Thinking



Level-3: Reveal a "secret" about NLP development

Prompting methods let out a secret how the technique of modern NLP progress

- Feature Engineering
- Architecture Engineering
- Objective Engineering
- Prompt Engineering

Feature Engineering

- Architecture Engineering
- Objective Engineering
- Prompt Engineering

- Paradigm: Fully Supervised
 Learning (Non-neural Network)
- Date: Before 2013
- Characteristic: Traditional machine learning model is mainly used, which requires manual feature definition of input text
- Typical Work:
 - CRF (Conditional Random Field)

- Feature Engineering
- Architecture Engineering
- Objective Engineering
- Prompt Engineering

- Paradigm: Fully Supervised Learning (Neural Network)
- **Date**: 2013 2018
- Characteristic:
 - Rely on neural networks
 - Do not need to manually define features, but should explore the network structure (e.g.: LSTM v.s CNN)
- Typical Work:
 - CNN for Text Classification

- Feature Engineering
- Architecture Engineering
- Objective Engineering
- Prompt Engineering

- **Paradigm**: Pre-train, Fine-tune
- **Date**: 2018-Now
- Characteristic:
 - context-dependent PLMs
 - Need to pay attention to the definition and selection of objective functions
- Typical Work: BERT

- Feature Engineering
- Architecture Engineering
- Objective Engineering
- Prompt Engineering

- Paradigm: Pre-train, Prompt, Predict
- **Date**: 2019-Now
- Characteristic:
 - NLP tasks are modeled entirely by relying on PLMs
 - More efforts on prompt design
- **Typical Work:** GPT3

PLMs and Downstream Tasks are Getting Closer and Closer



Secret let out from Prompt-based Learning

The history of modern natural language processing is essentially (probably) a history of changes in the relationship between downstream tasks and pre-trained language models (PLMs).



- (1) use pre-trained language models
- (2) use a better pre-trained language model
- (3) better use a pre-trained language model

Level-4: Beyond NLP

Prompting methods make

more modalities of signals (e.g. image)
 connected using natural language as relay
 node

New view for human to interact with data in the world



How does prompt-based research progress currently?

Website Resource for Prompt-based Research



Website Resource for Prompt-based Research

Timeline

Paperlist

Timeline of Prompt Learning

Q	Revisiting Self-Training for Few-Shot Learning of Language Model				
	Prompt-fix LM Tuning	04 October, 2021			
0	Towards Zero-Label Language Learning Tuning-free Prompting	19 September, 2021			
0	SentiPrompt: Sentiment Knowledge Enhanced Prompt-Tuning Sentiment Analysis Fixed-prompt LM Tuning	for Aspect-Based 17 September, 2021			
0	Reframing Instructional Prompts to GPTk's Language Tuning-free Prompting	16 September, 2021			
0	Language Models are Few-shot Multilingual Learners Tuning-free Prompting	16 September, 2021			
0	Can Language Models be Biomedical Knowledge Bases? Tuning-free Prompting; Fixed-LM Prompt Tuning	15 September, 2021			
0	Dialogue State Tracking with a Language Model using Schem Fixed-prompt LM Tuning	a-Driven Prompting 15 September, 2021			
0	Exploring Prompt-based Few-shot Learning for Grounded Dial Fixed-Prompt LM Tuning; Fixed-LM Prompt Tuning	og Generation 14 September, 2021			
0	Eliciting Knowledge from Language Models for Event Extraction Fixed-Prompt LM Tuning	on 11 September, 2021			
0	An Empirical Study of GPT-3 for Few-Shot Knowledge-Based \ Tuning-free Prompting	/QA 10 September, 2021			

Website Resource for Prompt-based Research

Paper Number

- Timeline
- Paperlist

ExplainaBoard - Prompt-based Learning

Full Survey can be obtained here.

Year	2018 2019 2020 2021 + more
Task	□ AR □ CKM □ CR □ CodeGen □ D2T □ EVALG □ FP + more
Pretrained LMs	ALBERT BART BERT CPM-2 CTRL Conv ELMo + more
Settings	Few Full Zero + more
Prompt Mining	automated hand-crafted + more

	Search						:			
	Year	Conf	Title 🖨		Task 🝦	PLMs 🛊	Citation	Bit		
0	2021	NAACL	It's Not Just Size That Matters: Small Language Models Are Also Few-Shot Learners Timo Schick,Hinrich Schütze		TC	ALBERT	58	Bib		
0	2021	ICLR	A Mathematical Exploration of Why Language Models Help Solve Downstream Tasks Nikunj Saunshi, Sadhika Malladi, Sanjeev Arora		Theory	GPT-2	5	Bit		
0	2021	ACL	Making Pre-trained Language Models Better Few-shot Learners Tianyu Gao,Adam Fisch,Danqi Chen		TC	RoBERTa	38	Bit		

TREND ANALYSIS

The number of papers over different tasks



How to apply prompting methods to diverse NLP tasks

□ More on classification/generation tasks + few-shot

□ Prediction-extensive tasks are under-explored

- How to apply prompting methods to diverse NLP tasks
- Tuning Strategy
 - \Box Head tuning
 - □ Adaptor tuning
 - □ Prompt tuning

- How to apply prompting methods to diverse NLP tasks
- Tuning Strategy
- Non-NLP Tasks
 - □ Multi-modal/Computer vision
 - \Box Biomedical

- How to apply prompting methods to diverse NLP tasks
- Tuning Strategy
- Non-NLP Tasks
- Annotation
 - □ Generate training samples
 - □ Annotate data

- How to apply prompting methods to diverse NLP tasks
- Tuning Strategy
- Non-NLP Tasks
- Annotation
- Pre-training

□ New pretraining framework

Summary of this talk

What is the "Prompt"?	 tool for human – PLM communication technique of making better use of pre- trained model by task reformulation 			
What is the general workflow? What are the design considerations?	 Prompt Construction Answer Construction Answer Prediction Answer-Label Mapping Pre-trained Model Choice Expanding the Paradigm Prompt-based Training Strate 			
What (unique) advantages?	 Level-1: Within Prompt Learning Level-2: Prompt Learning v.s. Fine-tuning Level-3: Modern NLP History 	More diverse NLP tasks		
How does prompt- based research progress?	Level-4: Beyond NLP	 Tuning Strategy Non-NLP Tasks Annotation Pre-training 		

Hopefully

