Prompt Learning Under the Large Language Model

Lili Sun
Silicon Lake College
Suzhou, China
113486860@qq.com

Zhenquan Shi*
Nantong University
NanTong, China
56254268@qq.com

***Abstract*-With the emergence of models such as chatGPT and Baidu AI Wenxin Yiyan , the research and application of NLP (Natural Language Processing) is increasingly centered on PLM (Pretrained Language Model),It marks that the current machine learning model has reached a new height. This article first introduces the background of the large language model, and introduces pre-training + fine-tuning and the current popular Prompt from the four major paradigms of NLP.Understand the workflow and function of Prompt , focusing on Prompt engineering, and structures, and looking ahead to future challenges for Prompt.**

Keywords-MLM,NLP , Prompt, chatGPT,PLM

I.THE WAVE OF LARGE LANGUAGE MODELS HAS ARRIVED

The rapid development of AI technology, especially breakthroughs in the field of natural language processing, has brought about many amazing progress. In the past few years, artificial intelligence technology has achieved from a simple question-answering robot to a more intelligent and flexible natural language processing model. And chatGPT is the representative of "quantitative change" caused by "quantitative change" of artificial intelligence technology, leading the leap-forward development in the field of artificial intelligence.

ChatGPT uses deep learning technology to automatically learn language rules and context information from a massive corpus, and store these information in a weight matrix. When the model uses these weights for natural language generation, it can generate more natural and fluent text according to the input information and contextual information.

ChatGPT3.5 was launched on November 30, 2022. Five days later, the number of users exceeded one million, and two months later, the number of monthly active users exceeded 100 million, making it the fastest growing consumer application in history.

ChatGPT adopts the route of "big data + large computing power + strong algorithm = large model" in the technical path, and explores a new paradigm in the direction of "basic large model + instruction fine-tuning", in which the basic large model is similar to the brain, and instruction fine-tuning is Interactive training, the combination of the two achieves language intelligence approaching to humans .

ChatGPT and Baidu AI Wenxinyiyan are inseparable from Prompt technology. Since the effect of the model largely depends on the quality and design of Prompt, Prompt has become a technical task . Whether chat GPT is a god or a demon depends on how your Prompt is . Prompt can not only affect the output quality of the model, but also improve the efficiency and scalability of the model .

Prompt is the abbreviation of "PRedictive OPTimization with Machine Learning", translated into Chinese as "Machine Learning Predictive Optimization". Prompt technology also becomes prompt learning , which usually limits the input of the artificial intelligence model to a specific range by converting the problem into a specific format of input, so that the machine can better understand the task , control the output of the model , and automatically generate human Linguistic text . Prompt technology has been widely used in search engines, social media and intelligent customer service and other fields . The advantage of Prompt is that it can reduce misunderstandings and mistakes caused by unclear language expression, so that it can perform specific tasks accurately and reliably.

II. THE 4TH PARADIGM OF NLP - PROMPT LEARNING

A. Origin of Prompt

By work such as chat GPT, many people began to explore how to obtain a relatively good model by modifying the downstream task to a language generation task when there is little or no training data . These works generally conform to the "4th paradigm" shown in Table 1, which is called hint learning .

Table 1 .Comparison of NLP paradigms

	Paradigm 1	Paradigm 2	3rd paradigm	4th paradigm
Paradigm	The era of probability and statistics	Deep learning era	The era of big language models	Tips for learning times
Principle	Non-neural network supervised learning	Representation learning + end-to-end classification	Pre-training + fine-tuning	Pretraining + hints + prediction paradigms
Represent	Bayes, svm, lr, etc.	Word2vec	Bert	Chatgpt
Task training data	A lot	A lot	A small amount	Little or no
Project	Feature engineering	Architecture engineering	Target project	Prompt project
Interpretability	Strongly explainable	Weakly explainable	Unexplainable	?

Hint learning achieves good task performance in zero- or few-shot scenarios by transforming downstream tasks,

2023 International Seminar on Computer Science and Engineering Technology (SCSET) | 979-8-3503-0147-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/SCSET58950.2023.00070

979-8-3503-0147-2/23/$31.00 ©2023 IEEE
DOI 10.1109/SCSET58950.2023.00070

adding expert knowledge, and fitting task inputs and outputs to the original language model . (Fig. 1) [1].

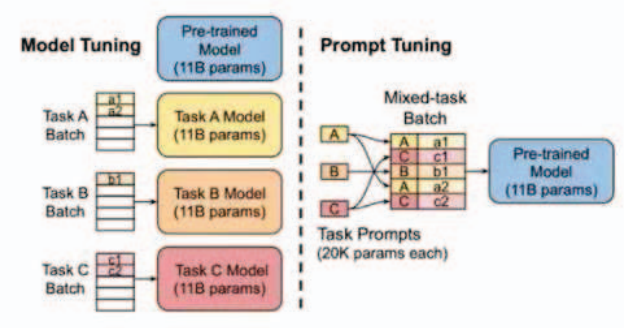


Figure 1. Schematic diagram of prompt learning

The pre-training + fine-tuning paradigm on the left side of Figure 1 . For different downstream tasks A, B, and C, the pre-training models with 11B magnitude parameters will be fine-tuned respectively to obtain three different 11B models after fine-tuning. The core is to adapt the pre-training models to downstream tasks. However, Prompt learning only uses the same pre-training model to build tasks for three different downstream tasks, eliminating the need for fine-tuning steps. The core is to adapt the downstream tasks to the pre-training model, so that you can make full use of the pre-trained pre-training model, greatly improving the efficiency of the pre-trained model.

B. Prompt Function

1) Setting Prompt Is Very Important for Generating High-Quality Text

For example, if we want chatGPT to generate a piece of technology news, we can give a prompt similar to "Please write a piece of news about artificial intelligence". This Prompt will help chatGPT better understand what we need and generate more text that meets the requirements.

2) Setting Prompt Can Also Help Us Control The Direction of Text Generated by ChatGPT

For example, if we want chatGPT to generate an article about travel for us, we can give a prompt of "Please write a travel guide". If we want chatGPT to generate an interesting story, we can give the prompt "Please write a thrilling story". In this way, we can control the style and content of the text generated by chatGPT to a certain extent , so as to get the results we want.

3) Setting Prompt Can also Help Us Improve the Interactive Ability of ChatGPT[2]

chatGPT to talk to us by setting Prompt . For example, we can give a prompt similar to "Who are you?", and then chatGPT will answer according to our question. Such an interactive process can increase the fun and playability of ChatGPT .

4) It Should Be Noted That When Setting the Prompt, You Need to Pay Attention to The Clarity and Accuracy Of the Prompt

If the prompt is not clear or accurate, ChatGPT may generate text that does not meet our requirements, or generate meaningless content. Therefore, when setting up the prompt, we need to carefully consider what kind of text we need, and then give the prompt as clear and accurate as possible.

In conclusion, setting the prompt is one of the key points of using ChatGPT . By setting the appropriate Prompt, we can help chatGPT generate high-quality text that meets the requirements, control the direction of the generated text, and improve the interactive ability of chatGPT. Therefore, when using chatGPT , we need to seriously consider the problem of setting Prompt.

III. TIPS TO LEARN BASIC PRINCIPLES

Currently, Prompt has four structures:

1) Prefix -Tuning

Add a series of continuous vectors (prefix prefix) before the output, keep the PLM parameters unchanged, and only train the vector. Mathematically, this involves optimizing the following log-likelihood objective given a trainable prefix matrix M_ϕ and a θ fixed PLM parameterized by :

$$\max_{\phi} \log P(y \mid x;\theta;\phi) = \max_{\phi} \sum_{y_i} \log P(y_i \mid h_{<i};\theta;\phi)$$

$h_{<i} = [h_{<i}^{(1)};\cdots;h_{<i}^{(n)}]$ the concatenation of all neural network layers at time step i . If the corresponding time step is within the prefix (h_i yes M_ϕ[i]), it is M_ϕ copied directly from , otherwise it is calculated using PLM.

The Prefix-Tuning structure is shown in Figure 2.[1]

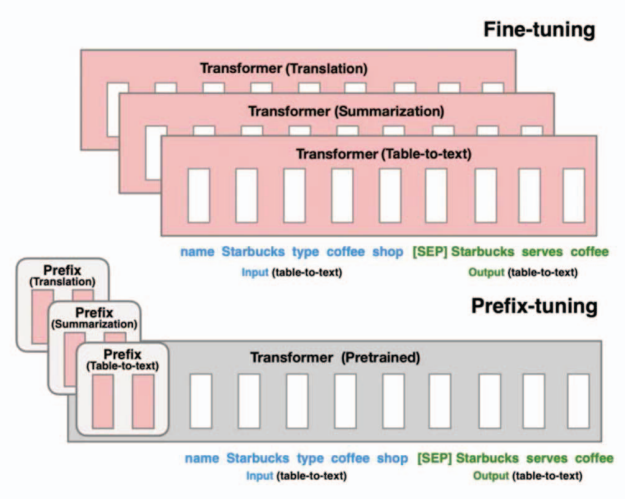


Figure 2. Prefix-Tuning Structure

2) Prompt-Tuning

The basic structure of Prompt-Tuning is shown in Figure 3.

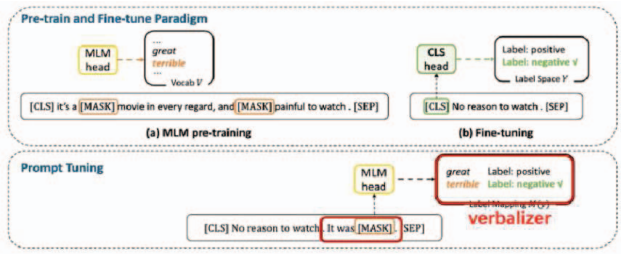


Figure 3. Basic Structure of Prompt-Tuning

Prompt-Tuning is to use the template to stimulate the factual knowledge in the pre-trained language template such as the BERT model, and then use the obtained factual knowledge as the answer space to map it to the target space to complete the required downstream tasks.

From an operational point of view, Prompt-Tuning usually inputs the original text to be processed together with a suggestive template (template project) into the pre-training template, and then the obtained answer (that is, the knowledge inspired by the template) is composed of The space is called the answer space), and the answer is mapped to the target label through the algorithm, as shown in the verbalizer in Figure 3.[3]

The composition of Prompt-Tuning is as follows:

Build templates. Through manual definition, automatic search, text generation and other methods, generate templates that should contain [MASK] tags related to a given sentence. For example, It was[MASK]. and spliced into the original text to get the input of Prompt-Tuning: [CLS]I like the Disney films very much.[SEP] It was[MASK].[SEP].

Tag word mapping. Because we are only interested in some words in the [MASK] part, we need to establish a mapping relationship. For example, if the word predicted by [MASK] is " great " , it is considered positive, and if it is "terrible " , it is considered negative .

objective function. If the input text is recorded as x, the prediction target is l (that is, label). Then in finetune, our goal is to calculate p(l|x), and in prompt learning, our goal is to calculate:

$$s_p(l \mid x) = M(v(l) \mid p(x))$$

Among them, $v(l)$ represents the word obtained through the mapping of "category->class tag word"; $p(x)$ represents the input obtained based on template transformation.

loss function. Note that L_{ce} *is the cross-* entropy loss of the classifier , and L_{mlm} is the loss when the language model predicts the mask, then the learning loss function is as follows:

$$L = (1-\alpha)L_{ce} + \alpha L_{mlm}$$

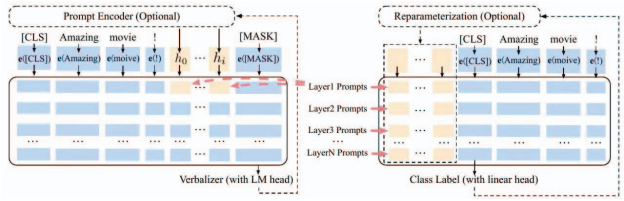
3) P-Tuning

P-Tuning is based on GPT-3, and the output of BiLSTM is used to represent Promptembedding, so that a certain degree of interaction can be generated inside Prompt. The basic structure is shown in Figure 4.

4) P-Tuning-v2

The previous several hint learning structures use only a frozen language model to fine-tune successive hints, greatly reducing the storage and memory usage per task during training, however, in the context of NLU, for normal-sized pre- trained models The performance is not ideal and cannot handle difficult sequence labeling tasks, indicating a lack of generalizability. P-Tuning-v2 can be generally effective across a wide range of model sizes and NLU tasks, with only 0.1%-3% fine-tuning parameters compared to earlier hint learning items, and is a strong baseline for future research.

Figure 4 is a comparison between P-Tuning and P-Tuning-v2:[4]



P-Tuning basic structure P-Tuning-v2 basic structure

Figure 4. Comparison Between P-Tuning and P-Tuning-v2

The embedding stored or calculated by the frozen pre-trained language model . Compared with P-Tuning, P-tuning v2 independently adds trainable continuous hints to the input of each transformer layer (like Prefix-Tuning). Furthermore, P-tuning v2 removes verbalizers with LM heads and returns to traditional category labels with plain linear heads to allow generalizability of its tasks.

IV. CONCLUSION

With the continuous development of natural language processing technology, we will encounter more and more vertical fields, and the speed of generating new vertical fields will also accelerate. This requires pre-trained language models to quickly form working capabilities under low-resource conditions. Prompt learning is another very good strategy after fine-tuning, and the future development prospects of Prompt engineers are very broad.

Hint learning is actually a kind of human-led analysis of the pre-trained language model, which can help us gain a deep understanding of the knowledge and capabilities of the pre-trained language model, which may lead to the emergence of a more powerful language model.

At present, most of the work using Prompt focuses on classification tasks and generation tasks, and there are fewer other tasks, because how to effectively link pre-training tasks and Prompt is still a problem worth exploring. In addition, the connection between the template and the answer has yet to be resolved. How to search at the same time or learn the best combination of the two is still very challenging.

Although the Prompt method has been successful in many cases, there are still few theoretical analyzes and guarantees for Prompt-based learning, making it difficult for people to understand why Prompt can achieve good results and why it has similar meanings in natural language . Prompt sometimes works very differently.

PLM has seen a lot of natural language in the human world during pre-training , so it is naturally influenced. For example, there are a lot of "The capital of China is "Beijing." in the training corpus, causing the model to think that it will predict "Beijing" when it sees "capital" next time, instead of focusing on which country's capital it is . In In the process of application, Prompt also exposed many other biases learned by PLM, such as racial discrimination, terrorism, gender antagonism, and so on.

REFERENCES

[1] Sylvestre-Alvise Rebuffi, Hakan Bilen, and Andrea Vedaldi, Learning multiple visual domains with residual adapters[J], In Advances in Neural Information Processing Systems, volume 30, pages 506–516.

[2] LUO Ge,ZHANG Xinpeng.Focusing on ChatGPT: development, impact, and challenges[J]. Chinese Journal of Nature.2023, 45(2):106-109.

[3] Liu Y L,Li H L,Bai X et al.A brief analysis of ChatGPT：historical evolution,current applications,and future prospects[J]. Journal of Image and Graphics,2023. 28(04):893-902.

[4] Xiao Liu, Kaixuan Ji, Yicheng Fu, Weng Lam Tam, Zhengxiao Du, Zhilin Yang, Jie Tang ,Prompt Tuning Can Be Comparable to Fine-tuning Universally Across Scales and Tasks [C] Proceedings of the 60th Annual Meeting of the Association of Computational Linguistics, 2022