

Parameter-Efficient Fine-Tuning (PEFT)


에이아이스쿨(AISchool) 대표
양진호 (솔라리스)


<http://aischool.ai>


<http://solarisailab.com>

Parameter-Efficient Fine-Tuning (PEFT)란?


- <https://github.com/huggingface/peft>

 README.md

 **PEFT**



State-of-the-art Parameter-Efficient Fine-Tuning (PEFT) methods



Parameter-Efficient Fine-Tuning (PEFT) methods enable efficient adaptation of pre-trained language models (PLMs) to various downstream applications without fine-tuning all the model's parameters. Fine-tuning large-scale PLMs is often prohibitively costly. In this regard, PEFT methods only fine-tune a small number of (extra) model parameters, thereby greatly decreasing the computational and storage costs. Recent State-of-the-Art PEFT techniques achieve performance comparable to that of full fine-tuning.

Seamlessly integrated with 🚀 Accelerate for large scale models leveraging DeepSpeed and Big Model Inference.

Supported methods:

1. LoRA: [LORA: LOW-RANK ADAPTATION OF LARGE LANGUAGE MODELS](#)
2. Prefix Tuning: [Prefix-Tuning: Optimizing Continuous Prompts for Generation](#), [P-Tuning v2: Prompt Tuning Can Be Comparable to Fine-tuning Universally Across Scales and Tasks](#)
3. P-Tuning: [GPT Understands, Too](#)

Parameter-Efficient Fine-Tuning (PEFT)란?

- Parameter-Efficient Fine-Tuning (PEFT) 방법은 모든 모델의 매개변수를 미세 조정(Fine-Tuning)하지 않고도 사전 훈련된 (pre-trained) 언어 모델(PLM)을 다양한 하위 작업에 효과적으로 적응시키는 데 도움을 줍니다.
- 대규모 PLMs의 미세 조정은 종종 비용이 많이 듭니다. 이러한 맥락에서 PEFT 방법은 소수의 (추가적인) 모델 매개변수만을 미세 조정하여 계산 및 저장 비용을 크게 줄입니다. 최근의 최첨단 PEFT 기술은 전체 미세 조정의 성능과 비교할 수 있는 성능을 달성합니다

Parameter-Efficient Fine-Tuning (PEFT)에서 지원하는 기법들

- PEFT에서 지원하는 모델들은 아래와 같습니다.
1. [LoRA](#) : LORA: LOW-RANK ADAPTATION OF LARGE LANGUAGE MODELS
 2. [Prefix Tuning](#) : Prefix-Tuning: Optimizing Continuous Prompts for Generation, P-Tuning v2: Prompt Tuning Can Be Comparable to Fine-tuning Universally Across Scales and Tasks
 3. [P-Tuning](#) : GPT Understands, Too
 4. [Prompt Tuning](#) : The Power of Scale for Parameter-Efficient Prompt Tuning
 5. [AdaLoRA](#) : Adaptive Budget Allocation for Parameter-Efficient Fine-Tuning
 6. [\(IA\)^3](#) : Few-Shot Parameter-Efficient Fine-Tuning is Better and Cheaper than In-Context Learning
 7. [MultiTask Prompt Tuning](#) : Multitask Prompt Tuning Enables Parameter-Efficient Transfer Learning