

适配器微调 (Adapter-tuning) 篇

来自: AiGC面试宝典



Just do it!

2024年05月19日 21:11



三

• 适配器微调 (Adapter-tuning) 篇

- 一、为什么 需要 适配器微调 (Adapter-tuning) ?
- 二、适配器微调 (Adapter-tuning) 思路?
- 三、适配器微调 (Adapter-tuning) 特点是什么?
- 四、AdapterFusion 思路 是什么?
- 五、AdapterDrop 思路 是什么?
- 六、AdapterDrop 特点 是什么?
- 七、MAM Adapter 思路 是什么?
- 八、MAM Adapter 特点 是什么?

一、为什么 需要 适配器微调 (Adapter-tuning) ?

1. 预训练模型参数量变多, 在特定任务下进行全量微调即昂贵又耗时;

二、适配器微调 (Adapter-tuning) 思路?

- 设计了 Adapter 结构 (首先是一个 down-project 层将高维度特征映射到低维特征, 然后过一个非线性层之后, 再用一个 up-project 结构将低维特征映射回原来的高维特征; 同时也设计了 skip-connection 结构, 确保了在最差的情况下能够退化为 identity), 并将其嵌入 Transformer 的结构里面;
- 在训练时, 固定住原来预训练模型的参数不变, 只对新增的 Adapter 结构进行微调。同时为了保证训练的高效性 (也就是尽可能少的引入更多参数)。

三、适配器微调 (Adapter-tuning) 特点是什么?

- 特点:
 - 通过在 Transformer 层中嵌入 Adapter 结构, 在推理时会额外增加推理时长。

四、AdapterFusion 思路 是什么?

- 思路: 一种融合多任务信息的 Adapter 的变体, 在 Adapter 的基础上进行优化, 通过将学习过程分为两阶段来提升下游任务表现。

五、AdapterDrop 思路 是什么？

- 思路：在不影响任务性能的情况下，对 Adapter 动态高效的移除，尽可能的减少模型的参数量，提高模型在反向传播（训练）和正向传播（推理）时的效率。

六、AdapterDrop 特点 是什么？

- 特点：
 - 通过从较低的 Transformer 层删除可变数量的 Adapter 来提升推理速度；
 - 当对多个任务执行推理时，动态地减少了运行时的计算开销，并在很大程度上保持了任务性能。

七、MAM Adapter 思路 是什么？

- 思路：一种在 Adapter、Prefix Tuning 和 LoRA 之间建立联系的统一方法。最终的模型 MAM Adapter 是用于 FFN 的并行 Adapter 和 软提示的组合。

八、MAM Adapter 特点 是什么？

- 特点：
 - 整体上来说，最终的模型 MAM Adapter 效果会优于单个高效微调方法。