每周研究进展阶段汇报

汇报人：杨凯冰
电　邮：tjuykb3022234232@163.com
时间段：2025 年 1 月 11 日 (周六) 至 2025 年 1 月 17 日 (周五)

一、本周工作：
1. 细读'Crab: A Unified Audio Visual Scene Understanding Model with Explicit cooperation'，并向冬暖学长回报，了解‘视音推理和生成’项目主要情况。
2. 复现基础 LoRA 代码
3. 学习数据标注知识，和文轩学长讨论数据集标注流程

二、思考总结：

Part 1.Crab model

论文主要分为两部分，一部分是数据集的构建，一部分是模型建立，模型建立中包括 Visual, Audio, Segmentation 部分和含有 interaction-aware LoRA 的 LLM 模型，并通过一系列实验验证该模型的效率和性能。论文创新点主要有两部分：

1. 新构造的数据集通过 MLLMs，将之前孤立的单词和词组进行组合，形成推理过程并纳入到新构造的数据集中，使得数据集涵盖各种乐器演奏的时间顺序、位置关系等，能够使模型清晰地理解任务之间的逻辑关联，从而在不同表征之间建立显式合作。

2. 设计 interaction-aware LoRA 结构，由 shared martix A 和多个 LoRA 头聚成，通过交互感知路由器结构 R 来动态调整每个 LoRA 头的权重，有效减少了视听数据异质性带来的干扰

Part 2.LoRA

对于任何一个矩阵 W 都可以对它进行低秩分解，把一个很大的矩阵分解为两个小矩阵 (A, B), 在训练过程中不去改变 W 的参数，而是去改变 AB [1]:

$$W_{new} = W_0 + W = W_0 + AB \tag{1}$$

最终在训练计算为

$$h = W_0 x + ABx = W_0 + \frac{\alpha}{r} ABx \tag{2}$$

$$s.t. \quad W_0 \in R^{n \times m}, A \in R^{n \times r}, B \in R^{r \times m} \tag{3}$$

其中 W_0 是原始矩阵 ($Embedding$ 等层中的 $weight$),W 是学习后更新的矩阵, 由低秩矩阵 A,B 的乘积得到;h 是输出；r 是低秩矩阵的秩,$r << n$ 且 $r << m$ 甚至 r 可以设置为 1；α 为可配置的超参数，与 r 共同构成缩放因子 α/r, 控制低秩矩阵对输出的影响。

创新点可以归结为如下内容：

- W 不是满秩的，含有较多冗余信息，则可以使用更加接近满秩的 A, B 代替，降低微调过程的参数
- 在训练过程中，引入 $lora\_A$, 和 $lora\_B$，并冻结了预训练的权重矩阵
- 在 $forward$ 中，根据 r 和 $merged$ 的值，动态决定是否使用 LoRA 调整卷积层的输出

通过这样微调后，一方面通过冻结预训练权重过于注重新数据集而泛化能力降低的风险，另一方面通过两个低秩矩阵降低调参规模，提高调参速率，并且能够作为即插即用的插件，插入各个训练层中使用。
其中较为重要的 $forward$ 部分代码添加个人注释后如下

```python
    def forward(self, x: torch.Tensor):
        if self.r > 0 and not self.merged:
            # 调用 nn.Embedding 的前向传播
            result = nn.Embedding.forward(self, x)
            # 使用 F.embedding 计算 A 矩阵对输入的映射
            after_A = F.embedding(
                x, self.lora_A.transpose(0, 1), self.padding_idx, self.max_norm,
                self.norm_type, self.scale_grad_by_freq, self.sparse
            )
            # 计算结果并添加 LoRA 权重的贡献
            result += (after_A @ self.lora_B.transpose(0, 1)) * self.scaling
            return result
        else:
            # 仅使用 nn.Embedding 的前向传播
            return nn.Embedding.forward(self, x)
```

但是 LoRA 存在可能的缺陷：

- 模型表达能力受限；由于知识对低秩矩阵的调整，无法做到对全参数调整的精细化，知识微调后性能提升不如全参数微调；
- 对超参数敏感；在使用 LoRA 微调的时候，需要尝试不同的 r 和，才能找到较为理想的超参数配置，带来较高的不确定性和调优难度；
- 知识遗忘风险；在低秩矩阵有较大改变的时候，可能会致使原来模型遗忘预训练学习到的知识。

Part 3. 数据集标注

在和文轩师兄对接后，学习数据集标准注意事项：

- 明确标注规则。制定统一的标注规则，如类别定义、边界划分、尺寸、流程等，同时提供相关示例进行说明；
- 一致性检查。出现不属于规定的数据 *unknown*，如何标注和处理，不会影响最后数据集的一致性；
- 数据集标准工具 labelme 的安装和使用

Part 3. 下周规划

1. 阅读大模型综述 *Visualunderstanding* 部分 [2]
2. 依据和文轩学长讨论得到的数据标注过程，进行数据标准
3. 重新阅读 “*Crab*” 论文

References

[1] E. J. Hu, Y. Shen, P. Wallis, Z. Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, “Lora: Low-rank adaptation of large language models,” *arXiv preprint arXiv:2106.09685*, 2021.

[2] Y. Z. e. a. Li C, Gan Z, “Multimodal foundation models: From specialists to general-purpose assistants,” *Foundations and Trends® in Computer Graphics and Vision*, 2024.