

课程作业

工程实践与科技创新IV-J课程作业

2024年4月26日

1. 介绍

本课程作业的目标是自行实现对简单大语言模型（如GPT-2）的微调训练以及更多细节上的探索。

2. 具体内容

2.1 第一部分：微调GPT-2

在第一部分，你需要实现对预训练模型的微调，这里我们**指定为GPT-2模型**。你大概率会使用到包括transformers, pytorch在内的python库，可能会用到类似GPT2Tokenizer.from_pretrained(“gpt2”)的命令。当遇到网络连接类似问题时，可通过将模型下载到本地解决。值得注意的是GPT-2有不同参数量的版本，这里**指定使用117M参数量的版本**。以下为可能用到的模型下载地址以及关键库transformers、pytorch使用说明的中英文文档以供查阅。另外，**建议善用搜索工具**寻找所需资料。

<https://hf-mirror.com/openai-community/gpt2/tree/main>

<https://huggingface.co/docs/transformers/index>

https://github.com/liuzard/transformers_zh_docs

<https://pytorch.org/docs/stable/index.html>

<https://pytorch-cn.readthedocs.io/zh/latest/>

这一部分的重点在于，你需要自行寻找合适的数据或公开数据集对模型微调并进行评估。值得注意的是，低参数量的GPT-2与现在的大语言模型相比性能相差很多，对有难度的任务测试结果不太好是正常的。评估形式比较自由，可以采用一些公开提出的指标分数，也可以人工评估或是借助模型（如ChatGPT）评估等，但是要反映出微调前后的对比情况。

注意：在一开始你应该先熟悉如何加载相应的模型，确定模型可以正常进行生成。同时理解其内部的各个组件对于之后的拓展探索和报告书写也有帮助。此外，允许在已有代码上进行修改，但请注

明来源。

2.2 第二部分：进一步的探索

在第二部分，你可以在以上基础下，选取**2~3**个方向进行探索，每一方向约占20%分数，量力而行，多多益善。旨在对原有模型的选择设置、数据或方法进行更大程度的调整。这里所使用的模型不再局限为GPT-2模型，以下是一些可供参考的探索方向：

- 实现不同的token编码方式；
- 实现不同的位置编码方式；
- 实现不同的注意力机制（线性注意力、Nyström注意力等）；
- 应用于中文数据；
- 应用于自己收集的私域数据；
- 模型、算法、评测基准等各层面的创新性改进；
- 大语言模型在新场景、数据集上的应用。

这里不对最后的性能做特别要求，更注重中间实现的过程与有效性的验证。若使用了其他模型，尤其是GitHub上其他人预训练了的模型作为基线，请注明来源。

3. 形式要求

本课程作业为小组合作形式，**2~3人一组**，最终以**书面报告+线下交流**的形式检验成果。

3.1 书面报告具体要求

报告分为**中期报告**和**最终报告**。

对于中期报告，需注意下面方面：

- a. 中期报告分数占10%
- b. 中期报告**截止时间：5月19日（第13周）**
- c. 你需保证第一部分的代码已能跑通，并能正常生成。
- d. 中期报告可以写关于你代码运行的README，并附上简要样例。
- e. 报告书写中英文皆可，最终与相应代码文件一起打包成压缩包进行提交。

对于最终报告，需注意下面方面：

- a. 最终报告分数占90%，第一部分50%，第二部分40%

- b. 最终报告**截止时间：6月9日（第16周）**
- c. 请**注明小组成员以及分工**，体现相应工作量，同等工作量下2人小组会比3人小组有酌情加分。
- d. 在第一部分任务中，尽量详细地说明模型的结构，生成原理以及前期调研验证的一些内容。
数据集方面更推荐选择一些有趣或针对某一指定任务的小型数据集，请简要介绍你所选择的数据集，并包含必要的数据处理说明。
必须有微调前、后的对比评估（评估方式可选特定指标、人工、ChatGPT等）。
- e. 在第二部分任务中，请尽量精简清晰地表达你的探索过程，不追求最终的性能提升，但请写明相应的原理、改进措施和实验结果等。
- f. 报告书写中英文皆可，最终与相应代码文件一起打包成压缩包进行提交。

3.2 线下交流具体要求

在提交了至少一版最终报告后，请私下联系助教预约时间见面交流（建议还是每周五下午上课时间）。助教会根据最终报告内容进行10分钟左右的提问交流。

最后请务必注意：

- 可以借鉴别人的代码，但请充分了解掌握最终提交的代码。在线下交流环节会对此进行考察，如发现其对于自己写的代码过于陌生，可能会扣除相应分数；
- 非特殊情况不允许迟交，否则没有线下交流，且本次作业只有及格分数。