人工智能 – LLM实践

小学数学应用题自动解题

复旦大学计算机科学技术学院

人工智能助教团队

2025-4-22

任务背景

赛题来源

- 小学数学应用题自动解题

小学数学应用题自动解题

中国计算机学会 & 题拍拍

智能算法　序列标注　教育

队伍 / 人数

2648 / 2955

比赛官网：https://www.datafountain.cn/competitions/467

任务背景

任务是什么

- 阅读理解是NLP中的一个常见任务，通常要求在大段文本中理解关键信息。数学应用题包含简单的文字表述，相对密集的推理和计算，是评估机器阅读理解能力的一个重要场景。同时，应用题也是K12教研的重要组成部分，如果机器能完美的理解题意，将会给AI在教育中的发展产生巨大的想象空间。

- 该任务是为了衡量现有机器学习模型在应用题理解方面的能力，模型读入一个应用题，输出该题的结果。为了降低任务的难度，赛题选择小学数学1-6年级校内题目。

任务背景

任务举例

1. Q: 商店有4框苹果，每框55千克，已经卖出135千克，还剩多少千克苹果?

   A: 85

2. Q: 玩具厂生产了960个电子玩具，每3个装一盒，每5盒装一箱，一共装了多少箱?

   A: 64

相当于给定问题，标签是对应的数字答案，类似于数学填空题

比赛数据

|____train.json（训练集）
|____test.json（测试集）
|____ submit.csv baseline提交结果文件示例

数据格式与特点

本赛题数据可直接使用整理好的数据，官方的数据有些乱，下面链接的数据进行了清洗

train.json中包含12000条训练数据; test.json中包含8000条测试数据

submit.csv 是baseline输出的结果，可作为提交模板进行参考

csv文件包含id和ret两列，其中id是test.json中的题目id，ret为预测结果

整理后的代码和数据：https://github.com/AI-FDU/Math_Solver

评估方法

提交说明

以csv文件格式提交结果到比赛平台，平台进行在线评分，实时排名。如果很久没有出分，请联系助教

提交入口在“作品提交”
每日每个账号可提交 3 次，对开多个账号提交答案不作限制，报告最后结果的用户名即可

评测标准

任务评估以正确率作为衡量标准：统计样本预测值与实际值一致的情况占整个样本的比例（衡量样本被正确标注的数量），即score = 正确数/总数，得分越高，成绩越好。

Baseline讲解：微调Qwen-0.5B

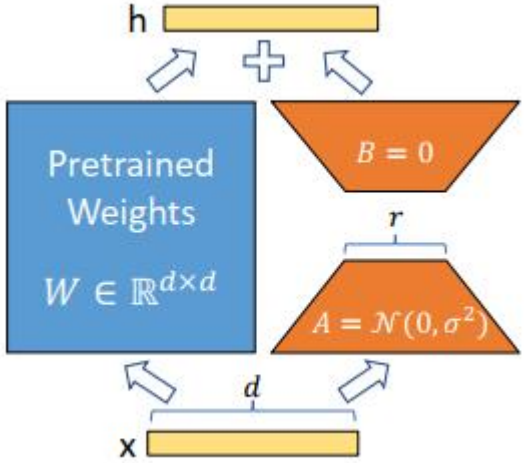
微调技术

如果把问题当成x，那么要输出的数学答案看成y，让模型进行SFT（有监督微调）即可

我们这里使用的基于LORA的PEFT方法进行微调：

PEFT（Parameter-Efficient Fine-Tuning）是 Hugging Face 提供的专门用于参数高效微调的工具库，外挂一个少量参数的可调小模型，无需微调原模型模型的参数， 显著降低训练成本。

LoRA（Low-Rank Adaptation）是 PEFT 支持的多种微调方法之一，简单理解一下，就是在模型的Linear层的旁边，增加一个“旁支”，这个“旁支”的作用，就是代替原有的参数矩阵W进行训练。



https://huggingface.co/Qwen/Qwen2.5-0.5B

Baseline讲解：微调Qwen-0.5B

微调技术

在代码层面要实现上述功能很简单：

```
from peft import get_peft_model, LoraConfig
from transformers import Trainer
model = …
config = LoraConfig(…)
model = get_peft_model(model, config)
trainer = Trainer(model=model, …. )
trainer.train()
```

即可开始训练

https://github.com/AI-FDU/Math_Solver/blob/main/qwen_ft.py

Baseline讲解

额外说明

- 该baseline只是出于大家了解整个任务流程的目的
- 本实践希望大家去尝试并跑通目前大模型训练的基本步骤
- 任务更多要求查看下面的可扩展方向

可扩展方向

方案1：思维链（COT）

- 不微调模型，直接使用它的能力，去优化prompt

- 标准的大模型输出，可以直接输出答案，简洁，但是在处理较数学问题时会影响其准确率
- Few-shot CoT模式下，提示中带有推理步骤的示例，通过示例引导模型学习推理模式
- Zero-shot CoT，通过在提示中加入特定关键词，如“Let's think step by step”，无需示例即可激活模型的推理能力

https://zhuanlan.zhihu.com/p/14949386849

可扩展方向

方案2：数据构建

- 当前的答案过于简单，没有解题的步骤，即思维链（COT）

  可利用更高级的大模型，将训练集的答案步骤补充出来：直接去问，把正确答案的步骤和错误答案的步骤分开来；把答案给大模型，让他补充出步骤；或者其他好的方法

- 当下数据存在不足的问题：可基于现有数据进行扩充，比如更改原题中的数字，合成新的数据

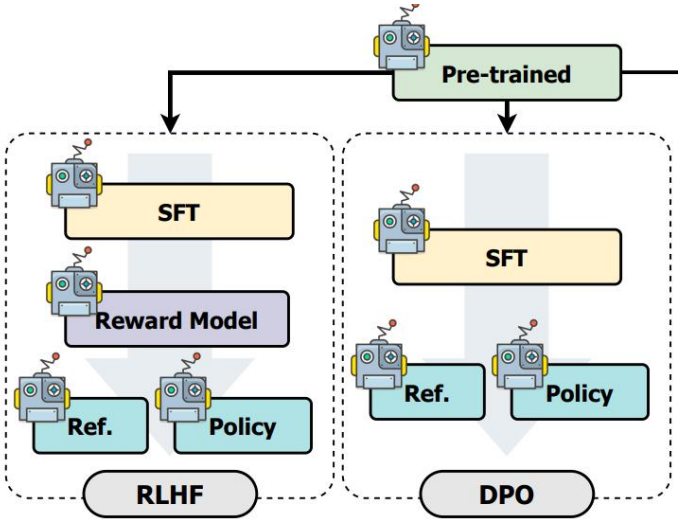
- 利用具有步骤的正确答案数据进行SFT

- 从含步骤的答案中提取出数字答案可使用正则表达式提取，或要求模型按照格式输出以及额外引入一个提取答案的模型

可扩展方向

方案3：RLHF+DPO

- RLHF需要的是偏好数据，之前构建的数据恰好可以得到正确答案的COT和错误答案的COT；给前者奖励，给后者惩罚
- 直接策略优化（DPO），最小化模型与最优策略之间对【优选-不优选】响应训练对的对数概率差异。



- 代码实现参考下面仓库，需构建DPOTrainer期望的格式

```
from trl import DPOConfig, DPOTrainer
```

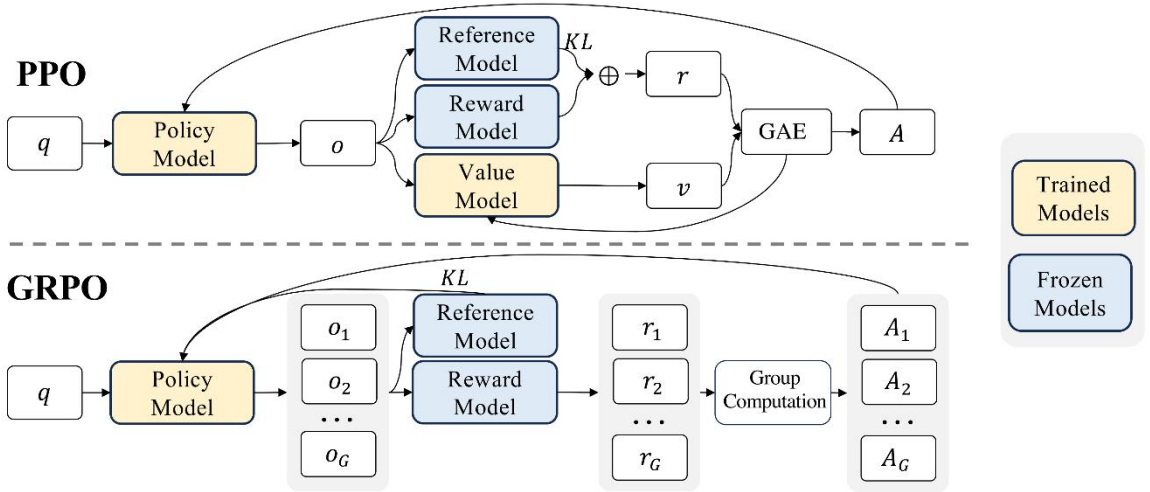
https://github.com/ShawhinT/YouTube-Blog/tree/main/LLMs/dpo

可扩展方向

方案4：GRPO

- 组相对策略优化（GRPO）是DeepSeek-R1模型中采用的一种创新的强化学习算法，旨在优化大型语言模型（LLMs）在复杂任务中的表现，如数学推理和代码生成。可参考Open-R1中的GRPO的复现代码。

生成一组响应：对于每个提示，从 LLM 中生成多个响应的一组。
对组进行打分（奖励模型）：获取组内所有响应的奖励分数。
计算组内相对优势：通过比较每个响应的奖励与组内平均奖励来计算优势。在组内对奖励进行归一化以得到优势。
优化策略：使用一个 PPO 风格的目标函数更新 LLM 的策略，但使用这些组内相对优势。



https://github.com/huggingface/open-r1

可扩展方向

方案说明

- 当前方案并不是所有，只是一些示例，鼓励去复现一些没有提到的方案
- 以上提到的都是大模型的解法，提供其他小模型的解法也能算一种方案
- 后面有提到按完成的方案来算分，方案的完成要有对应的实现、结果和实验分析
- 即使结果可能不好，但只要对方案做了好的尝试和探索，我们也是认可的

规则说明

总体规则

- 只允许对训练数据进行重新构建和增强，严禁处理测试数据
- 输出结果的推理模型限制为0.5B（推荐Qwen）或更小的模型（Bert等）
- 可以组队或是个人形式完成本次实践，个人会稍微减少要求
- 可以选取其他相同工作量的比赛或实践，但需要向助教报备

规则说明

■ 报名组队 & 个人完成

组队要求

- 按要求在腾讯文档进行组队，上限五人，不对人数做结果倾斜
  但组内摆烂者经认证后将对组内得分比例进行调整
- 中途不建议更换组队，极端情况除外
- 组队需要完成上述任务点，并于15周进行现有结果汇报
- 16周提交完整报告

个人要求

- 15周汇报前提交一份报告草稿，说明当前完成的内容，以防抄袭汇报团队的方案
- 16周提交完整报告

规则说明

比赛时间与计分

- 为减轻期末压力，本次课程比赛计分时间截止到15周汇报为止
- 16周主要是留出时间进行报告书写

- 比赛部分分数由数值分决定：

  假设得分是x，成绩计算为s1=x*15，刷满当然就满分
- 工作量分：完成的方案点数，团队组队每个点3分，个人完成每个点5分

  假设实现了t个方案，成绩计算为s2=t*3

总分：s=min(s1+s2, 15)

对于组队来说，即使比赛分数可能不高只有0.2，但尝试了4个方案依然可以拿满15分

个人的话，分数有0.3，尝试2个方案也能接近满分

实验要求

1. 比赛结果分：由分数决定- *15分*

2. 方法工作量和创新性：由报告和汇报体现- *12分*

最后得分为：min(15, s1+s2)，总分15分。

比赛截止日期: 2024年05月27日 !!
提交最好排名和分数截图、csv文件至elearning

报告截止日期: 2024年06月06日 !!
提交最终代码、方案报告（4页纸）、PPT（如有）至elearning

THANKS

人工智能助教团队

2025-4-22