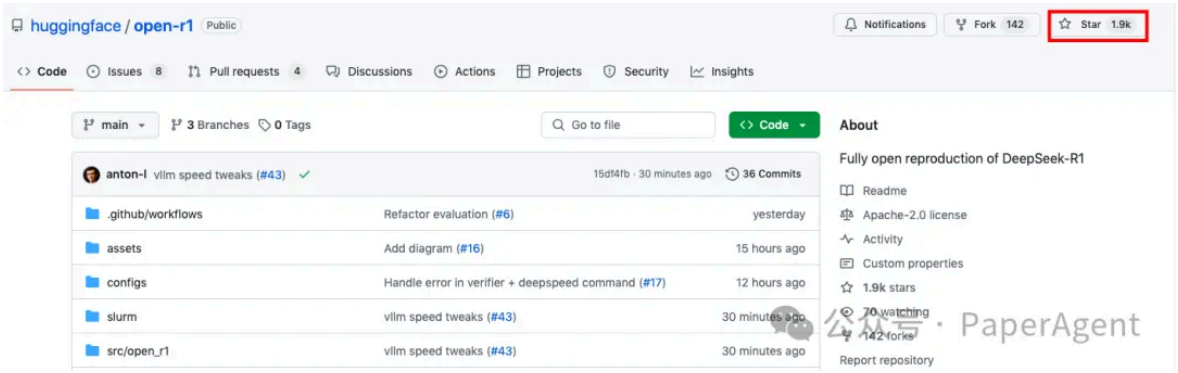
首个DeepSeek-R1全开源复现Open-R1来了

春城在下雪~ PaperAgent 2025年01月26日 09:56 云南

Open-R1：huggingface出品，**DeepSeek-R1的完全开源复现**，短短一天已经冲上1.9k Star，这个仓库仍在建设中。



Open-R1的目标是构建**DeepSeek-R1**流程中缺失的部分，以便每个人都可以复现并在此基础上进行开发。项目设计简单，主要包含以下内容：

- **src/open_r1** 包含用于训练和评估模型以及生成合成数据的脚本：
  - **grpo.py**：使用GRPO在给定数据集上训练模型。
  - **sft.py**：在数据集上对模型进行简单的SFT（监督微调）。
  - **evaluate.py**：在R1基准测试上评估模型。
  - **generate.py**：使用Distilabel从模型生成合成数据。
- **Makefile**：包含针对R1流程中每个步骤的易于运行的命令，利用上述脚本。

Open-R1将以**DeepSeek-R1**技术报告为指导，该报告大致可以分为三个主要步骤：

1. **第一步**：通过从DeepSeek-R1中提取高质量语料库，复现R1-Distill模型。
2. **第二步**：复现DeepSeek用于创建R1-Zero的纯强化学习（RL）流程。这可能涉及为数学、推理和代码创建新的大规模数据集。
3. **第三步**：展示能够通过多阶段训练从基础模型过渡到经过RL调整的模型。

训练模型

支持使用DDP（分布式数据并行）或DeepSpeed ZeRO-2和ZeRO-3来训练模型。要切换训练方法，只需更改configs文件夹中加速器（accelerate）YAML配置文件的路径即可。

以下训练命令是针对配备8块H100（80GB）显卡的单个节点配置的。如果使用不同的硬件或拓扑结构，可能需要调整批量大小和梯度累积步数。

- **SFT阶段**

```
accelerate launch --config_file=configs/zero3.yaml src/open_r1/sft.py \
    --model_name_or_path Qwen/Qwen2.5-Math-1.5B-Instruct \
    --dataset_name HuggingFaceH4/Bespoke-Stratos-17k \
    --learning_rate 2.0e-5 \
    --num_train_epochs 1 \
    --packing \
    --max_seq_length 4096 \
```

```
    --per_device_train_batch_size 4 \
    --per_device_eval_batch_size 4 \
    --gradient_accumulation_steps 4 \
    --gradient_checkpointing \
    --bf16 \
    --logging_steps 5 \
    --eval_strategy steps \
    --eval_steps 100 \
    --output_dir data/Qwen2.5-1.5B-Open-R1-Distill
```

- **GRPO**

```
accelerate launch --config_file configs/zero3.yaml src/open_r1/grpo.py \
    --output_dir DeepSeek-R1-Distill-Qwen-7B-GRPO \
    --model_name_or_path deepseek-ai/DeepSeek-R1-Distill-Qwen-7B \
    --dataset_name AI-MO/NuminaMath-TIR \
    --max_prompt_length 256 \
    --per_device_train_batch_size 1 \
    --gradient_accumulation_steps 16 \
    --logging_steps 10 \
    --bf16
```

数据生成

- **从一个小型蒸馏的R1模型生成数据**

1块H100显卡，从deepseek-ai/DeepSeek-R1-Distill-Qwen-7B生成数据

- **从DeepSeek-R1生成数据**

使用了2个节点，每个节点配备8块H100显卡，从DeepSeek-R1模型生成数据

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

推荐阅读
- 对齐LLM偏好的直接偏好优化方法：DPO、IPO、KTO
- 2024：ToB、Agent、多模态
- RAG全景图：从RAG启蒙到高级RAG之36技，再到终章Agentic RAG！
- Agent到多模态Agent再到多模态Multi-Agents系统的发展与案例讲解（1.2万字，20+文献，27张图）

欢迎关注我的公众号“**PaperAgent**”，每天一篇大模型（LLM）文章来锻炼我们的思维，简单的例子，不简单的方法，提升自己。

PaperAgent
日更，解读AI前沿技术热点Paper
215篇原创内容