DeepSeek R1 + unsloth + LoRA 大模型微调实战

1. 简介

一个使用 **unsloth** 训练框架和 **LoRA** 技术对 DeepSeek-R1-Distill-Qwen-7B 模型进行医疗领域微调实战，旨在提升模型在临床推理和诊断方面的能力。

```
# unsloth
https://github.com/unslothai/unsloth.git
```

2. 主要步骤

第1步：初始化设置

- 加载环境变量
- 登录 HuggingFace 和 Weights & Biases
- 初始化 wandb 项目用于实验追踪

```
# HuggingFace:
https://huggingface.co/settings/tokens

# Weights & Biases (wandb):
# 一个专门用于机器学习实验跟踪和可视化的工具
# 注册帐号
https://wandb.ai
# 获取token
https://docs.wandb.ai/quickstart/
```

第2步：模型加载

- 使用 unsloth 的 FastLanguageModel 加载预训练模型
- 采用 4bit 量化以节省显存
- 设置最大序列长度为 2048

```
# 模型: unsloth/DeepSeek-R1-Distill-Qwen-7B
https://huggingface.co/unsloth/DeepSeek-R1-Distill-Qwen-7B
```

第3步：提示模板设置和初始测试

- 定义医疗问答的提示模板
- 进行微调前的推理测试

唐国梁Tommy 原创出品 | © 2026 熵智未来（深圳）科技有限公司。保留所有权利。

第4步：数据集处理

- 定义训练用的提示模板
- 实现数据格式化函数
- 加载医疗推理数据集并进行格式化

```
# 数据集：FreedomIntelligence/medical-o1-reasoning-SFT

https://huggingface.co/datasets/FreedomIntelligence/medical-o1-reasoning-SFT
```

第5步：LoRA 配置

- 配置LoRA参数进行高效微调

第6步：配置训练参数和初始化训练器

- 使用 SFTTrainer 设置训练参数
- 配置批次大小、学习率、训练步数等

第7步：模型训练

第8步：微调后的模型推理测试

- 使用相同的测试用例评估微调后的模型效果

第9步：模型保存

- 保存完整模型
- 保存合并后的 16bit 模型

第10步：模型上传

- 将模型上传至 HuggingFace Hub

3. 微调环境配置

```
# 首先安装python3-venv
sudo apt install python3-venv

# 创建虚拟环境
```

唐国梁Tommy 原创出品 | © 2026 熵智未来（深圳）科技有限公司。保留所有权利。

```
python3 -m venv unsloth

# 激活虚拟环境
source unsloth/bin/activate

# 退出环境
# deactivate

# 依赖包安装

pip install unsloth

pip install wandb

pip install python-dotenv
```

4. 模型训练

```
# 直接执行脚本

python r1-finetuning-unsloth.py
```

1️⃣ 开始训练

```
(unsloth) (base) root@ubuntu:~/r1-finetuning#
(unsloth) (base) root@ubuntu:~/r1-finetuning# python r1-finetuning-unsloth.py
wandb: Appending key for api.wandb.ai to your netrc file: /root/.netrc
wandb: Currently logged in as: tommytang to https://api.wandb.ai. Use `wandb login --relogin` to force relogin
wandb: Using wandb-core as the SDK backend.  Please refer to https://wandb.me/wandb-core for more information.
wandb: Tracking run with wandb version 0.19.6
wandb: Run data is saved locally in /root/r1-finetuning/wandb/run-20250208_172802-ffvwa4wf
wandb: Run `wandb offline` to turn off syncing.
wandb: Syncing run robust-armadillo-1
wandb: ⭐ View project at https://wandb.ai/tommytang/Fine-tune-DeepSeek-R1-Distill-Qwen-7B%20on%20Medical%20COT%20Dataset
?apiKey=11a0ff012b65b101fdf6613d7c21f66a5960e623
wandb: 🚀 View run at https://wandb.ai/tommytang/Fine-tune-DeepSeek-R1-Distill-Qwen-7B%20on%20Medical%20COT%20Dataset/run
s/ffvwa4wf?apiKey=11a0ff012b65b101fdf6613d7c21f66a5960e623
wandb: WARNING Do NOT share these links with anyone. They can be used to claim your runs.
🦥 Unsloth: Will patch your computer to enable 2x faster free finetuning.
🦥 Unsloth Zoo will now patch everything to make training faster!
==((====))==  Unsloth 2025.2.5: Fast Qwen2 patching. Transformers: 4.48.3.
   \\   /|    GPU: NVIDIA GeForce RTX 4090. Max memory: 23.643 GB. Platform: Linux.
O^O/ \_/ \    Torch: 2.6.0+cu124. CUDA: 8.9. CUDA Toolkit: 12.4. Triton: 3.2.0
\        /    Bfloat16 = TRUE. FA [Xformers = 0.0.29.post2. FA2 = False]
 "-____-"     Free Apache license: http://github.com/unslothai/unsloth
Unsloth: Fast downloading is enabled - ignore downloading bars which are red colored!
Loading checkpoint shards: 100%|                                                | 2/2 [00:20<00:00, 10.34s/it]
```

2️⃣ 训练结束

唐国梁Tommy 原创出品 | © 2026 熵智未来（深圳）科技有限公司。保留所有权利。

```
Map: 100%|| 500/500 [00:00<00:00, 27561.10 examples/s]
Unsloth 2025.2.5 patched 28 layers with 28 QKV layers, 28 O layers and 28 MLP layers.
Map (num_proc=2): 100%|| 500/500 [00:00<00:00, 500.84 examples/s]
==((====))==  Unsloth - 2x faster free finetuning | Num GPUs = 1
   \\   /|    Num examples = 500 | Num Epochs = 1
O^O/ \_/ \    Batch size per device = 2 | Gradient Accumulation steps = 4
\        /    Total batch size = 8 | Total steps = 60
 "-____-"     Number of trainable parameters = 40,370,176
wandb: WARNING The `run_name` is currently set to the same value as `TrainingArguments.output_dir`. If this was not inten
ded, please specify a different run name by setting the `TrainingArguments.run_name` parameter.
{'loss': 2.7371, 'grad_norm': 0.316121369600296, 'learning_rate': 0.00018181818181818183, 'epoch': 0.16}
{'loss': 2.1612, 'grad_norm': 0.1971646547317505, 'learning_rate': 0.00014545454545454546, 'epoch': 0.32}
{'loss': 2.02, 'grad_norm': 0.21719777584075928, 'learning_rate': 0.00010909090909090909, 'epoch': 0.48}
{'loss': 1.9821, 'grad_norm': 0.2285522222518921, 'learning_rate': 7.272727272727273e-05, 'epoch': 0.64}
{'loss': 1.9342, 'grad_norm': 0.2156299501657486, 'learning_rate': 3.6363636363636364e-05, 'epoch': 0.8}
{'loss': 1.9135, 'grad_norm': 0.2369077056646347, 'learning_rate': 0.0, 'epoch': 0.96}
{'train_runtime': 157.5793, 'train_samples_per_second': 3.046, 'train_steps_per_second': 0.381, 'train_loss': 2.124693743
387858, 'epoch': 0.96}
100%|| 60/60 [02:37<00:00,  2.63s/it]
```

3 保存模型

```
Unsloth: Merging 4bit and LoRA weights to 16bit...
Unsloth: Will use up to 12.69 out of 28.5 RAM for saving.
Unsloth: Saving model... This might take 5 minutes ...
 71%|| 20/28 [00:00<00:00, 49.92it/s]
We will save to Disk and not RAM now.
100%|| 28/28 [00:03<00:00,  7.15it/s]
Unsloth: Saving tokenizer... Done.
Done.
README.md: 100%|| 627/627 [00:00<00:00, 3.81MB/s]
adapter_model.safetensors: 100%|| 162M/162M [01:10<00:00, 2.31MB/s]
Saved model to https://huggingface.co/tommysally/DeepSeek-R1-Medical-COT-Qwen-7B
tokenizer.json: 100%|| 11.4M/11.4M [00:05<00:00, 2.14MB/s]
Unsloth: You are pushing to hub, but you passed your HF username = tommysally.
We shall truncate tommysally/DeepSeek-R1-Medical-COT-Qwen-7B to DeepSeek-R1-Medical-COT-Qwen-7B
Unsloth: Merging 4bit and LoRA weights to 16bit...
Unsloth: Will use up to 12.7 out of 28.5 RAM for saving.
Unsloth: Saving model... This might take 5 minutes ...
100%|| 28/28 [00:03<00:00,  7.40it/s]
No files have been modified since last commit. Skipping to prevent empty commit.
Unsloth: Saving tokenizer... Done.
model-00001-of-00004.safetensors:   3%|| 132M/4.88G [02:18<1:17:52, 1.02MB/s]
model-00002-of-00004.safetensors:   3%|| 129M/4.93G [02:18<1:24:02, 952kB/s]
model-00003-of-00004.safetensors:   3%|| 140M/4.33G [02:18<1:00:58, 1.15MB/s]
Upload 4 LFS files:   0%|| 0/4 [00:00<?, ?it/s]
model-00004-of-00004.safetensors:  12%|| 126M/1.09G [02:18<18:47, 856kB/s]
```

4 上传模型

```
model-00002-of-00004.safetensors:  95%|| 4.69G/4.93G [1:06:50<
model-00002-of-00004.safetensors:  95%|| 4.69G/4.93G [1:06:50<
model-00002-of-00004.safetensors:  95%|| 4.69G/4.93G [1:06:50<
model-00002-of-00004.safetensors:  95%|| 4.69G/4.93G [1:06:51<
model-00002-of-00004.safetensors:  95%|| 4.69G/4.93G [1:06:51<
model-00002-of-00004.safetensors:  95%|| 4.69G/4.93G [1:06:51<
model-00002-of-00004.safetensors:  95%|| 4.69G/4.93G [1:06:51<
model-00002-of-00004.safetensors:  95%|| 4.69G/4.93G [1:06:51<
model-00002-of-00004.safetensors:  95%|| 4.69G/4.93G [1:06:52<
model-00002-of-00004.safetensors:  95%|| 4.70G/4.93G [1:06:52<
model-00002-of-00004.safetensors:  95%|| 4.70G/4.93G [1:06:52<
model-00002-of-00004.safetensors:  95%|| 4.70G/4.93G [1:06:52<
model-00002-of-00004.safetensors:  95%|| 4.70G/4.93G [1:06:52<
model-00002-of-00004.safetensors:  95%|| 4.70G/4.93G [1:06:53<
model-00002-of-00004.safetensors:  95%|| 4.70G/4.93G [1:06:53<
model-00002-of-00004.safetensors:  95%|| 4.70G/4.93G [1:06:53<
model-00002-of-00004.safetensors: 100%|| 4.93G/4.93G [1:08:20<00:00, 1.20MB/s]
Upload 4 LFS files: 100%|| 4/4 [1:08:21<00:00, 1025.39s/it]
Done.
Saved merged model to https://huggingface.co/tommysally/DeepSeek-R1-Medical-COT-Qwen-7B
wandb: 4 LFS files:  50%|| 2/4 [1:08:21<56:51, 1705.91s/it]
wandb: 🚀 View run robust-armadillo-1 at: https://wandb.ai/tommytang/Fine-tune-DeepSeek-R1-Distill-Qwen-7B%20on%20Medical%20
c21f66a5960e623
wandb: Find logs at: wandb/run-20250208_172802-ffvwa4wf/logs
```

5 模型已上传

唐国梁Tommy 原创出品 | © 2026 熵智未来（深圳）科技有限公司。保留所有权利。

Uploaded model

- **Developed by:** tommysally
- **License:** apache-2.0
- **Finetuned from model :** unsloth/deepseek-r1-distill-qwen-7b-unsloth-bnb-4bit

This qwen2 model was trained 2x faster with Unsloth and Huggingface's TRL library.

5. 源码解析

AI进阶学习

📺 **B站课堂**：https://space.bilibili.com/3546748815411393/pugv

🌐 **官方网站**：https://www.tgltommy.com/（国内访问需科学上网）

如果你正在关注大模型 Agent、强化学习后训练、RLHF、DPO、GRPO、RLVR 等前沿方向，欢迎关注我最新上线的精品课程：
《大模型 Agent 强化学习实战：从后训练、奖励设计到工业级对齐系统》

这门课程围绕当前大模型 Agent 强化学习的核心技术路线展开，不只是介绍零散算法和论文概念，而是希望帮助学习者建立一套完整的技术地图：从基础原理、论文脉络、关键算法，到开源项目源码解析与工业级系统实践，逐步理解 Agent RL 为什么重要、如何演进，以及在真实应用中如何落地。

课程配套专属飞书知识库 《Agent RL 学习宝典》，包含学习路线、论文资料、项目说明、源码阅读辅助资料和后续持续更新内容，方便长期查阅和跟进前沿技术。

Agent RL 变化很快，但越是变化快，越需要一套系统的学习框架。如果你希望系统进入 Agent RL 方向，而不是停留在碎片化了解阶段，这门课会是一个很好的起点。

课程官网：https://www.tgltommy.com/p/agent-rl（国内访问需科学上网）
B站课堂：https://www.bilibili.com/cheese/play/ss842375604

更多课程详情，扫描以下二维码咨询课程助理：

唐国梁Tommy 原创出品 | © 2026 熵智未来（深圳）科技有限公司。保留所有权利。

唐国梁Tommy 原创出品 | © 2026 熵智未来（深圳）科技有限公司。保留所有权利。

B站/公众号/YouTube：唐国梁Tommy

关注微信公众号
获取更多学习资源

官方网站：TGLTommy.com(需科学上网)

唐国梁Tommy 原创出品 | © 2026 熵智未来（深圳）科技有限公司。保留所有权利。