

- GPRO实验
 - 参数设定
 - 资源占用
 - 时间花费
 - 数据集

GPRO实验

参数设定

都是LoRA微调切Rank=8, lora_alpha=32

- Without unsloth

```
trainer = GRPOTrainer(model=model,
                      args=GRPOConfig(
                          output_dir="results/test_grpo",
                          learning_rate=1e-4,
                          num_train_epochs=5,
                          num_generations=4, # 越大越费时间
                          per_device_train_batch_size=2,
                          gradient_accumulation_steps=4,
                          max_grad_norm=1,
                          max_prompt_length=1024,
                          max_completion_length=512,
                          logging_dir="results/test_grpo/log",
                          logging_strategy=IntervalStrategy.STEPS,
                          logging_steps=50,
                          save_strategy=SaveStrategy.STEPS,
                          save_steps=2000,
                          lr_scheduler_type="cosine",
                          fp16=True,
                          do_train=True,
                      ),
                      reward_funcs=reward_func,
                      train_dataset=dataset_processed,
                      processing_class=tokenizer,
                      reward_processing_classes=tokenizer,
                      )
```

- With unsloth

```

trainer = GRPOTrainer(model=model,
args=GRPOConfig(
    output_dir="results/test_grpo_unsloth",
    learning_rate=1e-4,
    num_train_epochs=5,
    num_generations=4, # 越大越费时间
    per_device_train_batch_size=2,
    gradient_accumulation_steps=4,
    max_grad_norm=1,
    max_prompt_length=1024,
    max_completion_length=512,
    logging_dir="results/test_grpo_unsloth/log",
    logging_strategy=IntervalStrategy.STEPS,
    logging_steps=50,
    save_strategy=SaveStrategy.STEPS,
    save_steps=2000,
    lr_scheduler_type="cosine",
    bf16=True,
    do_train=True,
),
reward_funcs=reward_func,
train_dataset=dataset_processed,
processing_class=tokenizer,
reward_processing_classes=tokenizer,
)

```

资源占用

- Without unsloth

7	N/A	63C	P8	384M / 768M	40.53GiB / 79.65GiB	77%	Default	MEM: <div><div></div></div> 50.9%	UTIL: <div><div></div></div> 77%
---	-----	-----	----	-------------	---------------------	-----	---------	-----------------------------------	----------------------------------

- With unsloth

6	N/A	63C	P8	442M / 768M	10863MiB / 79.65GiB	99%	Default	MEM: <div><div></div></div> 13.3%	UTIL: <div><div></div></div> 99%
---	-----	-----	----	-------------	---------------------	-----	---------	-----------------------------------	----------------------------------

时间花费

- Without unsloth



- With unsloth



数据集

SG-KBQA项目中的WebQSP-GenDataset