强化学习实验报告 - PPO

生成时间：2025-05-22 01:28:08

实验设置

- 环境名：TimeLimit
- 智能体：PPO
- 训练轮数（episodes）：20
- 每轮最大步数（max_steps）：500

PPO算法超参数

- 学习率（lr）：0.0003
- 折扣因子（gamma）：0.99
- clip参数（clip_epsilon）：0.2
- 熵系数（entropy_coef）：0.05
- 是否使用 n-step TD（use_nstep_td）：False
- 网络结构（隐藏层维度）：128

实验结果

- 平均总回报：20.25
- 最佳回报：47.00（第 10 轮）
- 平均步数：20.25
- 奖励曲线图保存在：reward_curve.png
- 策略图保存在：policy_map.png

参数调整与实验观察

本次实验使用默认参数，尝试调整学习率、熵系数或 n-step 参数以优化效果。

--- 实验报告自动生成完毕 ---