

强化学习实验报告

生成时间: 2025-05-27 23:19:17

实验设置

- 环境名: TimeLimit
- 智能体: DeepQNetwork
- 训练轮数 (episodes) : 100
- 每轮最大步数 (max_steps) : 500

实验结果

- 平均总回报: 10.69
- 最佳回报: 20.00 (第 62 轮)
- 平均步数: 5.11
- 回报图保存在: reward_curve.png
- Q 表策略图保存在: q_table_policy.png

参数调整与实验观察

本次实验使用默认参数, 可尝试调整学习率或 epsilon 策略改善效果。

--- 实验报告自动生成完毕 ---