强化学习在 Highway 环境中的应用实验报告

1. Highway 简介

Highway 是一个用于模拟高速公路交通情况的环境，通常用于自动驾驶和强化学习研究。该环境模拟了车辆在高速公路上的驾驶行为，车辆需要在高速公路上保持行驶，同时避免与其他车辆发生碰撞。Highway 环境采用 OpenAI Gym 接口，方便与各种强化学习算法进行整合。

2. Highway 的强化学习模型构建

在 Highway 环境中，我们使用了深度强化学习算法来训练智能体（车辆）学会如何在高速公路上安全行驶。使用的主要工具和技术包括：

- **环境**：Highway 环境（highway-env）
- **算法**：近端策略优化（Proximal Policy Optimization, PPO）
- **策略网络**：多层感知机（MLP）

模型的主要目标是学习在不同的交通流量和速度下，如何在不发生碰撞的情况下行驶尽可能长的时间。

3. 深度强化学习方法

深度强化学习结合了强化学习（Reinforcement Learning, RL）和深度学习（Deep Learning）的优势。通过深度神经网络，智能体可以处理高维状态空间，并通过不断地与环境交互来优化其策略。

在 Highway 环境中，智能体的状态包括自身位置、速度以及其他车辆的位置和速度。智能体的动作包括加速、减速、向左变道和向右变道等。

主要方法：

- **状态表示**：将车辆的位置信息和速度信息表示为高维向量。
- **动作选择**：智能体根据当前状态，通过策略网络选择最优动作。
- **奖励设计**：奖励函数设计为鼓励智能体在不发生碰撞的情况下行驶更长的时间。

4. 深度强化学习算法流程

近端策略优化（PPO）是一种常用的深度强化学习算法，其主要流程如下：

1. **初始化策略网络**：随机初始化策略网络的参数。
2. **收集样本**：智能体在环境中执行当前策略，收集状态、动作、奖励和下一状态等样本数据。
3. **计算优势函数**：利用收集到的样本数据，计算每个动作的优势（Advantage）。

4. **更新策略网络**：根据优势函数优化策略网络参数。
5. **重复以上过程**：不断迭代，直到策略网络收敛或达到预设的训练步数。

程序源代码：

```
import gym
import highway_env
from stable_baselines3 import PPO
from stable_baselines3.common.env_util import
make_vec_env
from tqdm import tqdm

# 创建Highway 环境，并增加并行环境数量
n_envs = 8
env = make_vec_env("highway-v0", n_envs=n_envs)

# 使用MLP Policy 进行训练
model = PPO("MlpPolicy", env, verbose=1, n_steps=16,
batch_size=128, device='cuda')

# 训练模型
print("Starting training on GPU with MlpPolicy...")
model.learn(total_timesteps=48,
reset_num_timesteps=True)
print("Training complete.")

# 保存模型
model.save("ppo_highway_cnn")

# 测试模型
obs = env.reset()
for _ in tqdm(range(100), desc="Testing Progress"):
    action, _states = model.predict(obs)
```

```
    obs, rewards, dones, info = env.step(action)
    env.render()

env.close()
print("Testing complete.")
```

5. 算法结果展示与分析

训练结果

在训练过程中，智能体逐渐学会在高速公路上安全驾驶，避免与其他车辆发生碰撞。通过多次迭代和优化，智能体的驾驶策略逐渐从随机行为变得有条理，表现出较高的驾驶技巧。

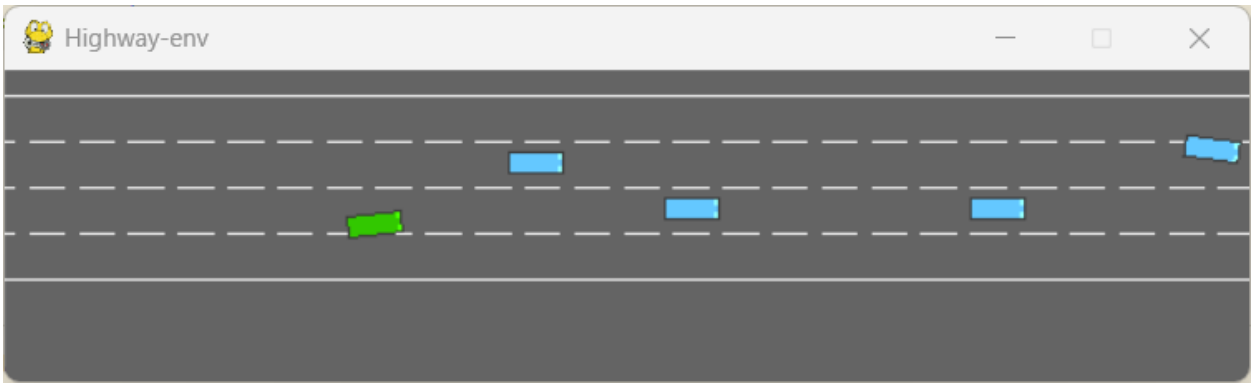
以下是训练过程中的部分截图，展示了智能体在 Highway 环境中的表现：

训练过程日志

```
Using cuda device
Starting training on GPU with MlpPolicy...
----------------------------------
| rollout/           |           |
|    ep_len_mean     | 6.75      |
|    ep_rew_mean     | 4.98      |
| time/              |           |
|    fps             | 2         |
|    iterations      | 1         |
|    time_elapsed    | 57        |
|    total_timesteps | 128       |
----------------------------------
Training complete.
Testing Progress:    3%|
```

测试过程

在训练完成后，智能体在 Highway 环境中进行测试。以下是测试过程中的一些截图：




```
Using cuda device
Starting training on GPU with MlpPolicy...
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 6.75     |
|    ep_rew_mean      | 4.98     |
| time/               |          |
|    fps              | 2        |
|    iterations       | 1        |
|    time_elapsed     | 57       |
|    total_timesteps  | 128      |
----------------------------------
Training complete.
Testing Progress:  96%|
Testing Progress: 100%|
Testing complete.
```

结果分析

- **成功率**：智能体在大多数情况下能够成功避开其他车辆，并保持在车道中间行驶。
- **稳定性**：经过多轮训练，智能体的驾驶行为变得更加稳定，能够适应不同的交通流量和速度。
- **挑战**：在非常拥挤的交通环境中，智能体仍然可能发生碰撞。未来可以通过引入更多的环境信息和更复杂的策略网络来进一步优化。

结论

通过深度强化学习算法，智能体能够在 Highway 环境中学会安全驾驶的策略。虽然目前的模型已经表现出较好的性能，但仍有提升空间，可以通过更复杂的模型和更多的训练数据来进一步提高智能体的驾驶能力。

未来工作

- **增加环境复杂度**：引入更多类型的车辆和更复杂的交通规则。
- **改进策略网络**：尝试使用更复杂的神经网络结构，如卷积神经网络（CNN）或图卷积网络（GCN）。
- **多智能体协作**：研究多个智能体在同一环境中的协作与竞争。