Project题目：GridWorld环境下的Model-Free Prediction方法对比研究

任务目标：
在一个自定义的GridWorld环境中，实现并对比以下四种Model-Free Prediction算法：

1. **Monte Carlo (MC) Prediction**
2. **Temporal Difference (TD(0)) Prediction**
3. **Forward View TD(λ) Prediction**
4. **Backward View TD(λ) with Eligibility Traces**

通过实验分析不同方法在策略评估（Policy Evaluation）中的性能差异。

项目要求

1. 环境构建

- 设计一个5×5的GridWorld，包含：
  - 普通格子（奖励=0）
  - 陷阱格子（奖励=-10，终止状态）
  - 目标格子（奖励=+10，终止状态）
  - 随机策略（如每个动作概率=0.25）
- 可自定义障碍物位置或随机生成地图。

2. 算法实现

- **Monte Carlo**：
  - 基于完整Episode更新状态价值函数。
  - 实现首次访问(First-Visit)和每次访问(Every-Visit)两种模式。
- **TD(0)**：
  - 单步更新，对比不同学习率(α)的影响。
- **Forward View TD(λ)**：
  - 实现λ-return的计算，对比λ=0, 0.5, 1的效果。
- **Backward View TD(λ)**：
  - 使用资格迹（Eligibility Traces）。

3. 实验与分析

- **收敛速度**：绘制各算法下状态价值函数的收敛曲线（如RMSE随Episode的变化）。

- **参数敏感性**：
  - 分析λ对TD(λ)算法的影响。
  - 对比MC和TD在不同Episode数量下的表现。
- **偏差-方差权衡**：定性讨论MC（高方差、无偏）与TD（低方差、有偏）的差异。
- 对Forward/Backward View TD(λ)，分别测试**λ=0**、**0.5**、**1**时的表现：
  - 绘制RMSE随Episode变化的收敛曲线（对比两种实现）
  - 记录算法运行时间，分析计算效率差异
  - 可视化λ=0.5时的资格迹累积过程（仅Backward View）
  - 在相同λ值下，对比Forward和Backward View的最终价值函数差异
- 尝试实现**动态λ调整策略**（如随时间衰减），并进行相应分析。

交付内容

1. **代码**：
   - 模块化的Python实现（如分 `environment.py` 、 `algorithms.py` 、 `visualization.py` ）。
   - 提供Jupyter Notebook示例调用代码。
2. **报告**：
   - 问题描述。
   - 方法与算法核心代码说明。
   - 实验结果图表与分析讨论。
   - 超参数设置（α, γ, λ等）的分析讨论。

评分标准

项目	分值
环境与算法正确性	30%
实验设计完整性	10%
代码规范与可读性	10%
实验报告	50%