**Домашнее задание 1: Выбор задачи и подготовка среды**

**Цели:**

1. Определить задачу, подходящую для методов обучения с подкреплением.
2. Подготовить и протестировать выбранную среду для экспериментов.

**Описание:**

1. **Выбор задачи**

**Доступные среды:** рассмотрим OpenAI Gym, Unity ML-Agents и собственные задачи:

- **OpenAI Gym**: подходит для быстрого прототипирования RL-задач. Множество готовых сред, таких как "CartPole", "MountainCar", "Atari", “ Taxi-v3”.
- **Unity ML-Agents**: используется для сложных визуальных задач, требующих 3D-симуляции.
- **Собственные задачи**: актуально для специфических исследований или узких применений.

**Описание задачи:** в среде "Taxi-v3" агент управляет такси в сетке 5x5. Цель — забрать пассажира в указанной точке и доставить его в целевую точку, избегая штрафов.

**Цели задачи:**

- Минимизировать общее время выполнения задачи.
- Избегать штрафов, таких как неправильные действия (высадка пассажира не в целевой точке).

**Обоснование выбора:**

- Задача "Taxi-v3" имеет конечное число состояний (500) и действий (6), что делает её удобной для изучения базовых алгоритмов обучения с подкреплением
- RL позволяет обучиться оптимальной стратегии на основе проб и ошибок, минимизируя штрафы и время доставки.
- Среда наглядна, проста для визуализации и не требует значительных вычислительных ресурсов.

2. **Подготовка среды**

1) **Установка необходимого программного обеспечения:**

```
pip install  gym
```

2) **Структура среды:**

- **Состояния**:
    - Положение такси на сетке (25 возможных позиций).
    - Текущее положение пассажира (5 возможных состояний: 4 фиксированные точки + в такси).
    - Целевая точка доставки (4 фиксированные точки).

- Общее количество состояний: = 500

- **Действия (6 вариантов)**:
  - Движение на север.
  - Движение на юг.
  - Движение на восток.
  - Движение на запад.
  - Посадка пассажира.
  - Высадка пассажира.

- **Награды**:
  - +20: за успешную доставку пассажира.
  - −1: за каждое движение (штраф за время).
  - −10: за некорректные действия (например, высадка пассажира не в целевой точке).

**3) Тестирование среды:**

Скрипт для проверки работоспособности Taxi-v3:

```
import gym
import numpy as np
np.bool8 = bool #для корректной синхронизации версий

def test_environment():
    #создаем среду
    env = gym.make("Taxi-v3", render_mode="ansi").env
    print("Среда успешно создана!")

    #проверяем пространство состояний и действий
    print(f"Количество состояний: {env.observation_space.n}")
    print(f"Количество действий: {env.action_space.n}")

    #сбрасываем среду и визуализируем начальное состояние
    initial_state = env.reset()[0]
    print("Начальное состояние:")
    print(env.render())

    #выполняем несколько случайных действий
    for i in range(5):
        action = env.action_space.sample()
        next_state, reward, terminated, truncated, info = env.step(action)
        done = terminated or truncated

        print(f"\nШаг {i + 1}:")
        print(f"Действие: {action}")
        print(f"Новое состояние: {next_state}")
```

```
        print(f"Награда: {reward}")
        print(f"Эпизод завершён: {done}")
        print(env.render())

        if done:
            print("Эпизод завершён. Сброс среды.")
            env.reset()

    print("Тестирование среды завершено.")

test_environment()
```

### 3. Результаты

```
Среда успешно создана!
Количество состояний: 500
Количество действий: 6
Начальное состояние:
+---------+
|R: | : :G|
| : | : : |
| : : : : |
|▒| : | : |
|Y| : |B: |
+---------+


Шаг 1:
Действие: 1
Новое состояние: 206
Награда: -1
Эпизод завершён: False
+---------+
|R: | : :G|
| : | : : |
|▒: : : : |
| | : | : |
|Y| : |B: |
+---------+
  (North)


Шаг 2:
Действие: 1
Новое состояние: 106
Награда: -1
Эпизод завершён: False
+---------+
|R: | : :G|
|▒: | : : |
| : : : : |
| | : | : |
|Y| : |B: |
+---------+
  (North)
```

```
Шаг 3:
Действие: 3
Новое состояние: 106
Награда: -1
Эпизод завершён: False
+---------+
|R: | : :G|
|▒: | : : |
| : : : : |
| | : | : |
|Y| : |B: |
+---------+
  (West)


Шаг 4:
Действие: 1
Новое состояние: 6
Награда: -1
Эпизод завершён: False
+---------+
|▒: | : :G|
| : | : : |
| : : : : |
| | : | : |
|Y| : |B: |
+---------+
  (North)


Шаг 5:
Действие: 5
Новое состояние: 6
Награда: -10
Эпизод завершён: False
+---------+
|▒: | : :G|
| : | : : |
| : : : : |
| | : | : |
|Y| : |B: |
+---------+
  (Dropoff)

Тестирование среды завершено.
```