Московский государственный технический университет

им. Н.Э. Баумана

Факультет «Информатика и системы управления»

Кафедра «Системы обработки информации и управления»



**Лабораторная работа № 5**

**По курсу «методы машинного обучения в АСОИУ»**

**«Обучение на основе временных различий»**

**Выполнил:**

студент ИУ5-24М

Ширшов А.С.

**Проверил:**

Гапанюк Ю.Е.

Подпись:

29.02.2024

Москва, 2024

**Задание**

На основе рассмотренного на лекции примера реализуйте следующие алгоритмы:

* SARSA
* Q-обучение
* Двойное Q-обучение

Для любой среды обучения с подкреплением (кроме рассмотренной на лекции среды Toy Text / Frozen Lake) из библиотеки Gym (или аналогичной библиотеки).

**Ход работы**

Возьмем за основу базовый пример из лекции. Код на рисунке ниже.

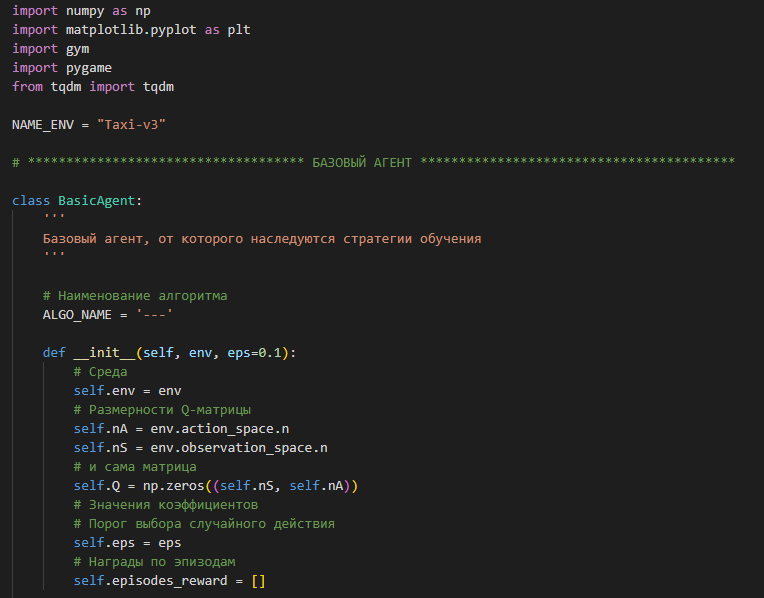


Рисунок 1 - Описание базового агента

В качестве среды возьмём taxi-v3.

**Описание среды Taxi-v3**

**Сеточный мир:**

Среда представляет собой квадратную сетку размером 5×5.

В сетке находятся несколько фиксированных точек: начальное положение пассажира, его цель и несколько препятствий.

**Действия агента:**

Агент может выполнять одно из 6 действий:

Движение на юг (0)

Движение на север (1)

Движение на восток (2)

Движение на запад (3)

Поднять пассажира (4)

Высадить пассажира (5)

**Состояния:**

Состояние среды определяется позицией такси, позицией пассажира и целью.

Всего имеется 500500 состояний (25 позиций такси × 5 возможных позиций пассажира × 4 возможных целей).

**Вознаграждение:**

Агент получает награду -1 за каждое перемещение.

Награда -10, если агент пытается поднять пассажира в неправильной позиции или высадить его в неправильном месте.

Награда +20 за успешную высадку пассажира в нужной позиции.

**Цель:**

Перевести пассажира с начальной позиции к цели с минимальными затратами времени и энергии.

На рисунках ниже представлены части реализации SARSA, Q-обучение, Двойное Q-обучение.

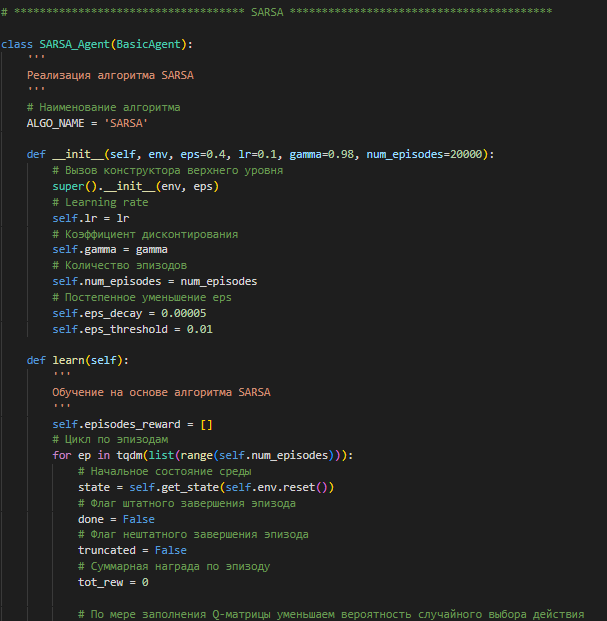


Рисунок 2 - Часть реализации Sarsa

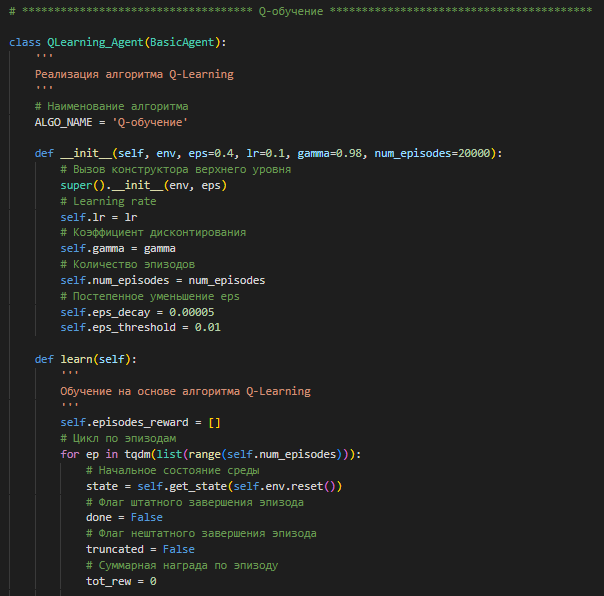


Рисунок 3 - Часть реализации Q-Learning

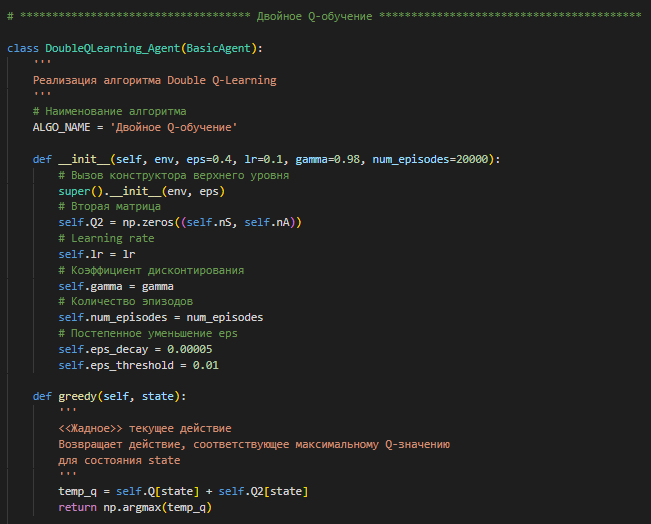


Рисунок 4 - Часть реализации двойного Q-обучения

Примеры работы алгоритмов. Запускается небольшая игра, где можем наблюдать процесс.

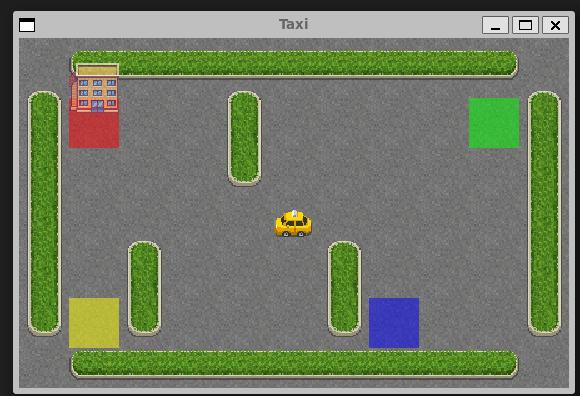


Рисунок 5 - Работа Q-Learning

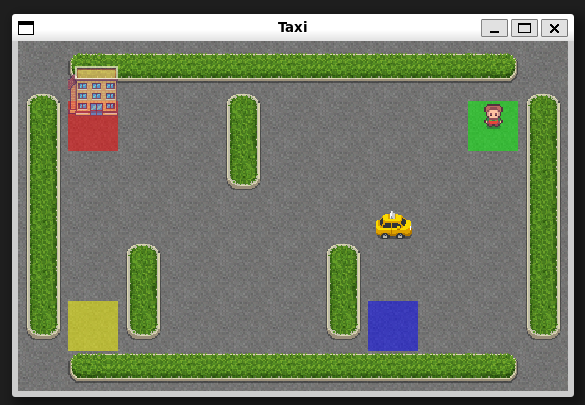


Рисунок 6 - Работа Sarsa

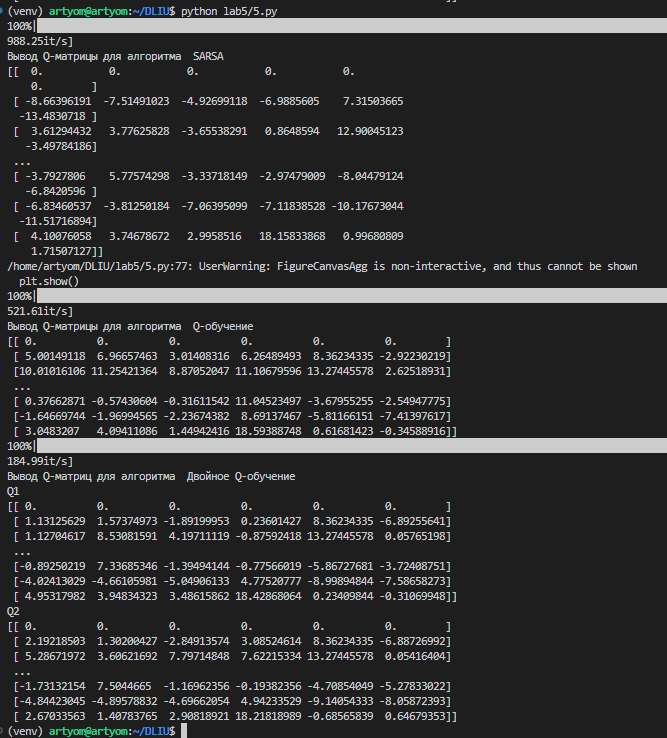


Рисунок 7 - Вывод матриц алгоритмов

**Вывод**

* Алгоритмы SARSA, Q-обучения и двойного Q-обучения могут быть успешно применены для обучения агентов в различных средах обучения с подкреплением.
* Выбор параметров алгоритмов (коэффициент обучения, коэффициент дисконтирования, эпсилон-жадная стратегия) может существенно влиять на результаты обучения.
* Применение двойного Q-обучения может улучшить результаты обучения по сравнению с обычным Q-обучением.