МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра МОЭВМ

ОТЧЕТ
по лабораторной работе №1
по дисциплине «Обучение с подкреплением»
Тема: Реализация DQN для среды CartPole-v1

Студент гр. 0306 _____ Сизов А.Р.

Преподаватель _____ Глазунов С.А.

Санкт-Петербург
2025 г.

Цель работы.

Реализация DQN для среды CartPole-v1. Исследование влияния различных параметров: архитектура сети, значения gamma и epsilon_decay, влияние epsilon на скорость обучения

Задание.

1. Реализация DQN
2. Измените архитектуру нейросети (например, добавьте слои).
3. Попробуйте разные значения gamma и epsilon_decay.
4. Проведите исследование как изначальное значение epsilon влияет на скорость обучения

Выполнение работы.

1. Реализация DQN

Основная цель агента — максимизировать суммарную награду, удерживая шест в вертикальном положении как можно дольше. В качестве Q-функции применялась полносвязная нейронная сеть с несколькими скрытыми слоями. Базовая архитектура включала входной слой: принимает 4 параметра состояния среды (позиция и скорость тележки, угол и угловая скорость шеста). Агент обучался в течение 600 эпизодов, каждый из которых длился до 500 шагов или до падения шеста.

2. Изменение архитектуры нейронной сети.

Рассмотрим зависимости эффективностей от типа сети.

Малая сеть (expanded): Обучение идет медленно, максимальная производительность не достигнута даже за 600 эпизодов из-за недостаточной емкости модели для сложных данных.

Средняя сеть (balanced): Демонстрирует сбалансированное обучение, до ~270 шагов эффективность на уровне малой сети. Однако далее, 300 - 600 наблюдается сильный рост эффективности и стабильной работы.

Большая сеть (expanded): Показывает быстрый начальный рост и наилучшие

результаты

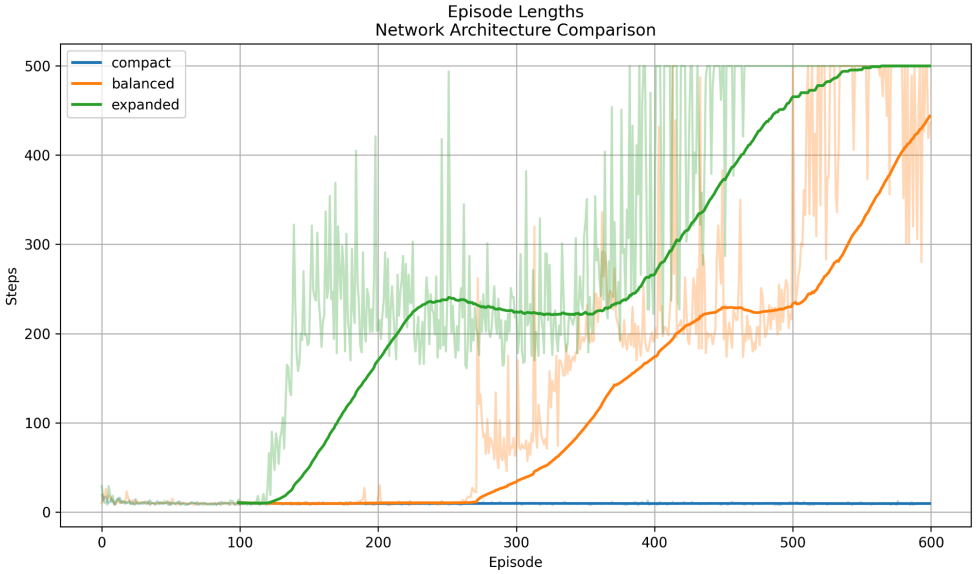

Рис 1. Графики эффективностей сети в зависимости от архитектуры

3. Изменение параметра gamma.

Коэффициент gamma отвечает за интерес к долгосрочной выгоде

При рассмотрении результатов, можно прийти к выводу, о том, что лучшая эффективность сети достигается при наибольшем значении параметра. Однако до 400 эпизодов разницы между 0.9 и 0.999 практически не наблюдается, т.к скорости роста эффективности схожи. После 400 эпизодов при наименьших параметрах показателя наблюдается явное превосходства у наибольших значений.

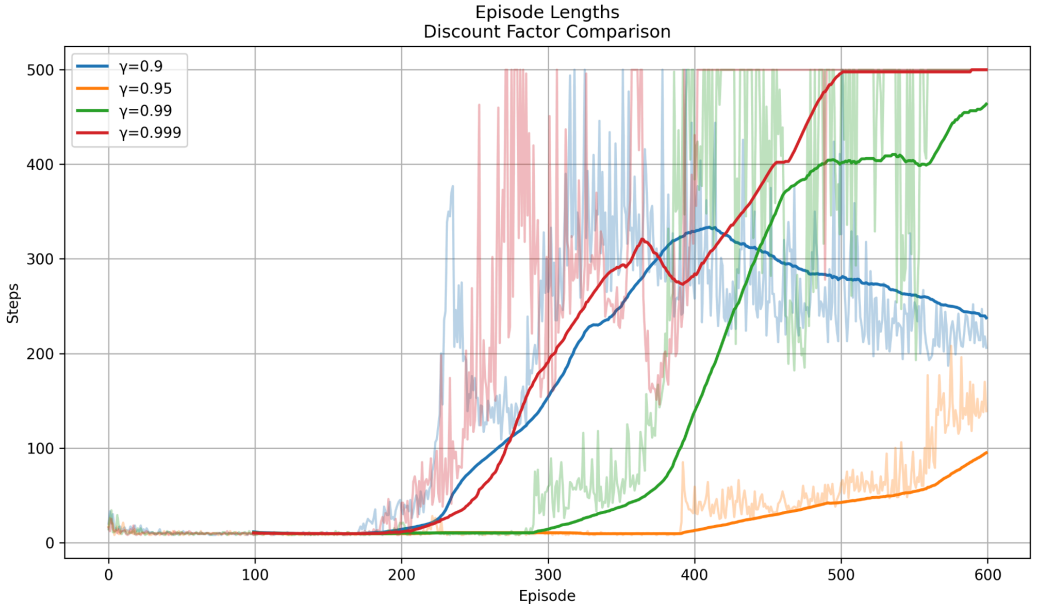

Рис 2. Графики эффективностей сети в зависимости от цены награды.

4. Изменение параметров epsilon_decay и start.

Параметры start и epsilon_decay влияют на частоту выбора случайных действий и то, как эта частота будет уменьшаться.
Наилучшие результаты показывают параметры start = 0.9 и decay = 0.95, что обуславливается созданием условий, при которых модель получает достаточно “случайного” опыта, и при этом может эффективно его использовать.

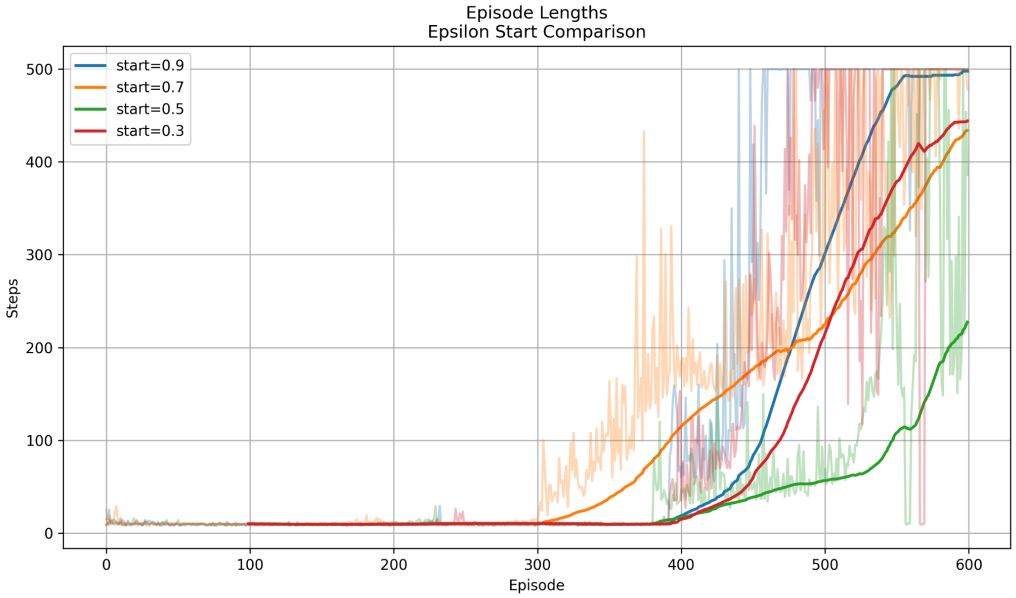

Рис 3. График эффективностей сети в зависимости от параметра start.

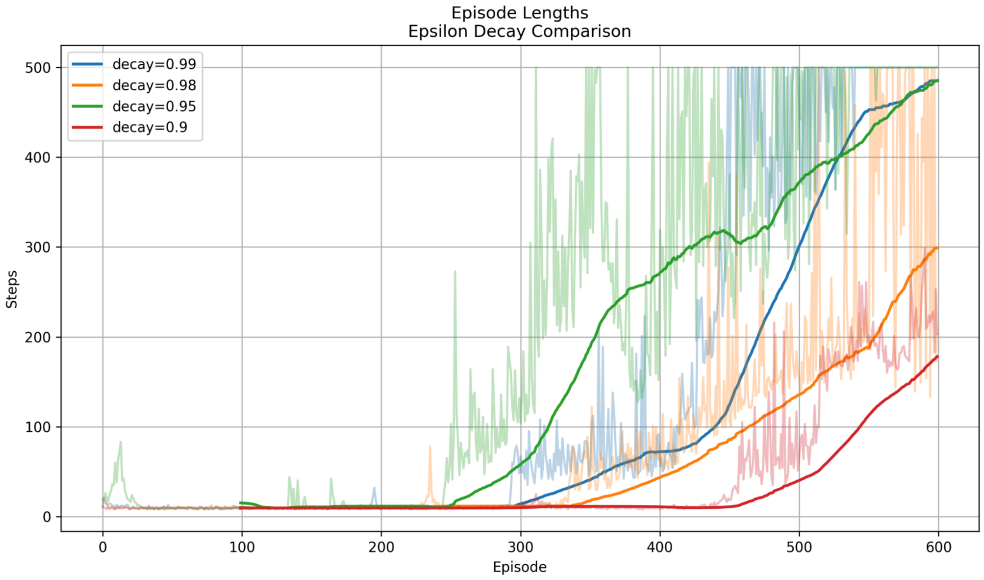

Рис 4. График эффективностей сети в зависимости от параметра decay.

Выводы.

Была выполнена реализация DQN для среды CartPole-v1, осуществлена его практическая реализация с использованием фреймворка PyTorch. Было проведено исследование влияния изменения некоторых параметров сети на ее результат.