|  |  |
| --- | --- |
| Gerb-BMSTU_01 | **Министерство науки и высшего образования Российской Федерации Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет имени Н.Э. Баумана**  **(национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана)** |

# Факультет «Информатика и системы управления» Кафедра ИУ5 «Системы обработки информации и управления»

Курс «Методы машинного обучения» Отчет по рубежному контролю №2

Выполнил: студент группы ИУ5-22М

Ульбашев А.Н.

20.05.2023

Проверил:

|  |
| --- |
| преподаватель каф. ИУ5 |
| Гапанюк Ю.Е. |

Москва, 2023 г.

# Задание:

Для одного из алгоритмов временных различий, реализованных Вами в соответствующей лабораторная работе:

* SARSA
* Q-обучение
* Двойное Q-обучение

Осуществите подбор гиперпараметров. Критерием оптимизации должна являться суммарная награда.

# Ход работы

Для проведения работы была выбрана среда обучения с подкреплением CliffWalking из библиотеки Gym.

Проведем подбор гиперпараметров для алгоритма SARSA. Критерий оптимизации

– суммарная награда. В класс SARSA\_Agent добавлен метод sum\_rewards, подсчитывающий итоговую суммарную награду:

def sum\_rewards(self): # Суммарная награда

sum\_rewards = sum(self.episodes\_reward) print('Cуммарная награда SARSA: ', sum\_rewards)

Изначальные гиперпараметры: eps=0.4, lr=0.1, gamma=0.98, num\_episodes=20000. Результаты изменения суммарной награды от скорости обучения lr представлены в

Таблице 1. График зависимости суммарной награды от lr представлен на рис.1.

Таблица 1. Результаты изменения суммарной награды от скорости обучения lr

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Суммарная награда | Награда  за последний эпизод | lr | eps | gamma | episodes |
| -529192 | -17 | 0,025 | 0,4 | 0,98 | 20000 |
| -499310 | -17 | 0,05 | 0,4 | 0,98 | 20000 |
| -484401 | -17 | 0,1 | 0,4 | 0,98 | 20000 |
| -488352 | -17 | 0,15 | 0,4 | 0,98 | 20000 |
| -488035 | -19 | 0,2 | 0,4 | 0,98 | 20000 |
| -497641 | -17 | 0,25 | 0,4 | 0,98 | 20000 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 0 0, | 05 0 | 1 0, | 15 0 | 2 0, | 25 0 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

Рис.1. График зависимости суммарной награды от lr Результаты изменения суммарной награды от параметра eps представлены в



Суммарная награда SARSA от lr

-480000

-485000

-490000

-495000

-500000

-505000

-510000

-515000

-520000

-525000

-530000

-535000

,

,

,3

lr

Суммарная награда

Таблице 2. График зависимости суммарной награды от параметра eps представлен на рис.2.

Таблица 2. Результаты изменения суммарной награды от параметра eps

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Суммарная награда | Награда  за последний эпизод | eps | lr | gamma | episodes |
| -321854 | -15 | 0,015 | 0,1 | 0,98 | 20000 |
| -322896 | -15 | 0,025 | 0,1 | 0,98 | 20000 |
| -322810 | -15 | 0,05 | 0,1 | 0,98 | 20000 |
| -329450 | -15 | 0,1 | 0,1 | 0,98 | 20000 |
| -351044 | -15 | 0,2 | 0,1 | 0,98 | 20000 |
| -422912 | -17 | 0,3 | 0,1 | 0,98 | 20000 |
| -484401 | -17 | 0,4 | 0,1 | 0,98 | 20000 |
| -586762 | -17 | 0,5 | 0,1 | 0,98 | 20000 |
| -806757 | -17 | 0,6 | 0,1 | 0,98 | 20000 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 0 0 | 1 0 | 2 0 | 3 0 | 4 0 | 5 0 | 6 0 |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Рис.2. График зависимости суммарной награды от параметра eps Результаты изменения суммарной награды от параметра gamma представлены в



Суммарная награда SARSA от eps

0

,

,

,

,

,

,

,7

-100000

-200000

-300000

-400000

-500000

-600000

-700000

-800000

-900000

eps

Суммарная награда

Таблице 3. График зависимости суммарной награды от параметра gamma представлен на рис.3.

Таблица 3. Результаты изменения суммарной награды от параметра gamma

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Суммарная награда | Награда  за последний эпизод | gamma | lr | eps | episodes |
| -321112 | -15 | 0,96 | 0,1 | 0,015 | 20000 |
| -319923 | -15 | 0,97 | 0,1 | 0,015 | 20000 |
| -321854 | -15 | 0,98 | 0,1 | 0,015 | 20000 |
| -321521 | -15 | 0,99 | 0,1 | 0,015 | 20000 |
| -321573 | -15 | 0,995 | 0,1 | 0,015 | 20000 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 55 0, | 96 0,9 | 65 0, | 97 0,9 | 75 0, | 98 0,9 | 85 0, | 99 0,9 | 95 |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

Рис.3. График зависимости суммарной награды от параметра gamma



Суммарная награда SARSA от gamma

-319500

0,9

1

-320000

-320500

-321000

-321500

-322000

gamma

Суммарная награда

**Вывод:** в результате подбора гиперпараметров лучшими значениями оказались: eps=0.015, lr=0.1, gamma=0.97, num\_episodes=20000. При этом, при уменьшении eps стратегия агента приближалась к максимальной ( к движению по краю обрыва).