

Zadanie 6 - Raport

Piotr Niedziałek

Czerwiec 2024

1 Wstęp

Celem laboratorium była implementacja algorytmu uczenia ze wzmocnieniem Q-Learningu oraz przeprowadzenie badań hiperparametrów. W badanym algorytmie agent aktualizuje wartości funkcji jakości dla par stan-akcja, odwiedzając kolejne stany i otrzymując za to odpowiednie nagrody. Agent wybiera najlepszą akcję zgodnie ze swoją aktualną wiedzą, lub poruszając się losowo. Prawdopodobieństwo z jakim agent będzie eksplorował lub eksploatował środowisko definiowane jest zgodnie ze strategią epsilon zachłanną lub Boltzmanna.

2 Implementacja

Zadanie zaimplementowano w pythonie. Skorzystano z bibliotek: gym, numpy, matplotlib, os, pandas. Struktura kodu:

- `train_agent` - aktualizowanie wartości tablicy Q
- `evaluate` - obliczanie średnich wartości nagrody i odchylenia standardowego dla 5 odpaleń algorytmu dla różnych ziaren
- `plot_results` - narysowanie średniej wartości nagrody w zależności od epizodu uczenia
- `calculate_statistics` - obliczenie średniej wartości, odchylenia standardowego i ilości udanych prób dla 10 ostatnich epizodów
- główna pętla programu - wyznaczanie wykresów, obliczenia dla kolejnych rodzajów i wartości hiperparametrów

3 Badania

Eksperymenty zostały przeprowadzone w środowisku "Frozen Lake" z biblioteki gym na planszy o wielkości 8 na 8 z włączonymi poślizgami. Badano hiperparametr α - współczynnik uczenia, ϵ - parametr eksploracji, γ - współczynnik dyskontowania, τ - parametr strategii Boltzmanna.

- α : [0.1, 0.3, 0.5, 0.7, 0.9]
- ϵ : [0.5, 0.7, 0.9, 0.95, 0.99]
- γ : [0.2, 0.4, 0.6, 0.8, 1.0]
- τ : [0.1, 0.5, 1.0, 2.0, 5.0]

4 Wyniki eksperymentów

	param_name	param_value	avg_last_10	std_dev_last_10	successful_episodes
0	alpha	0.100000	0.860000	0.156205	10
1	alpha	0.300000	0.540000	0.180000	9
2	alpha	0.500000	0.740000	0.128062	10
3	alpha	0.700000	0.240000	0.080000	2
4	alpha	0.900000	0.120000	0.097980	0

	param_name	param_value	avg_last_10	std_dev_last_10	successful_episodes
10	epsilon	0.200000	0.760000	0.149666	10
11	epsilon	0.400000	0.660000	0.156205	10
12	epsilon	0.600000	0.740000	0.128062	10
13	epsilon	0.800000	0.620000	0.188680	9
14	epsilon	1.000000	0.780000	0.208806	9

	param_name	param_value	avg_last_10	std_dev_last_10	successful_episodes
5	gamma	0.500000	0.240000	0.120000	3
6	gamma	0.700000	0.720000	0.097980	10
7	gamma	0.900000	0.880000	0.132665	10
8	gamma	0.950000	0.580000	0.188680	9
9	gamma	0.990000	0.740000	0.128062	10

	param_name	param_value	avg_last_10	std_dev_last_10	successful_episodes
15	tau	0.100000	0.020000	0.060000	0
16	tau	0.500000	0.020000	0.060000	0
17	tau	1.000000	0.020000	0.060000	0
18	tau	2.000000	0.020000	0.060000	0
19	tau	5.000000	0.020000	0.060000	0

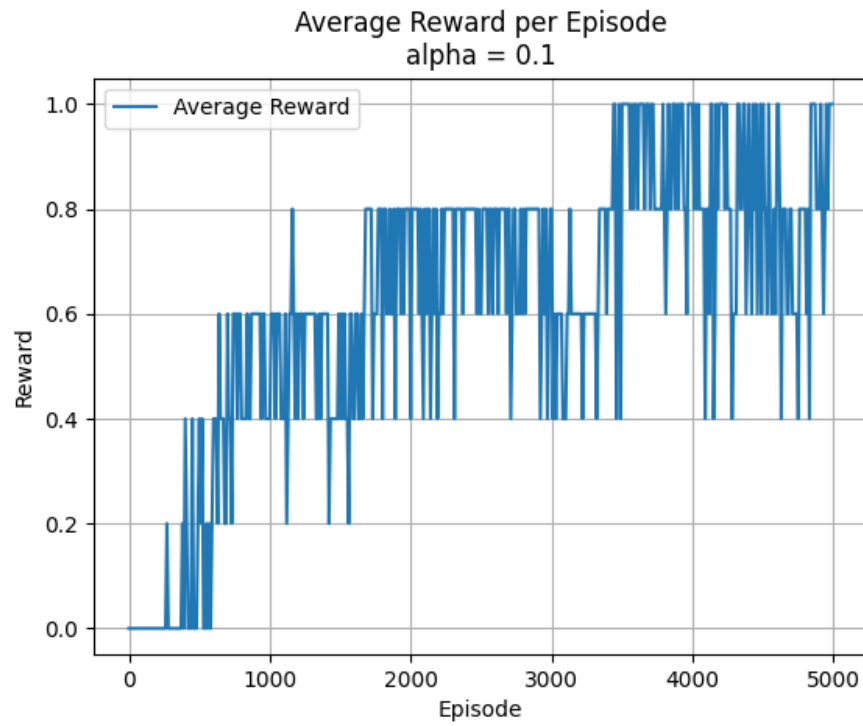


Figure 1: $\alpha = 0.2$

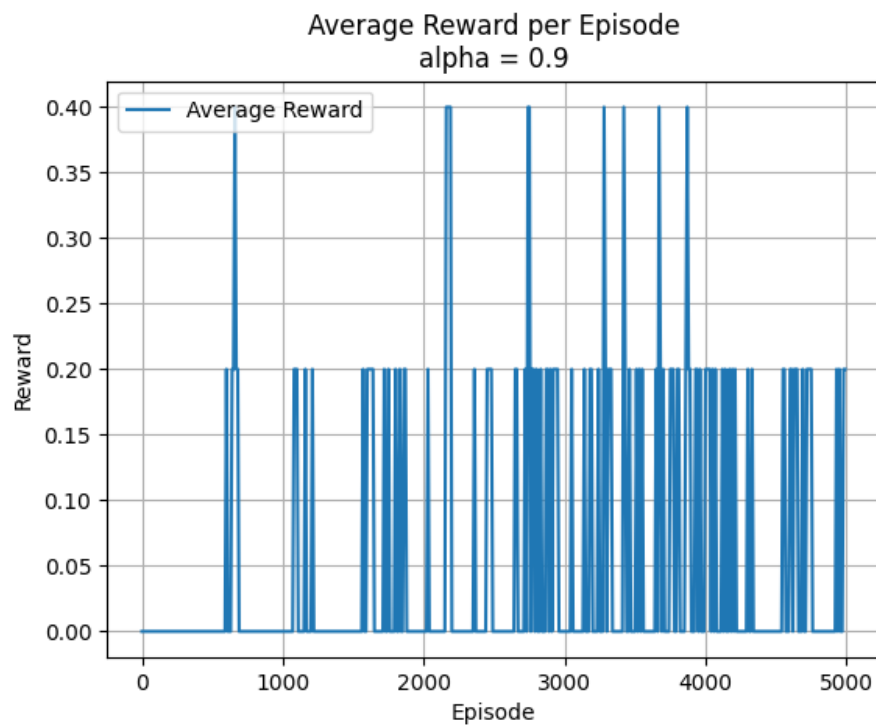


Figure 2: $\alpha = 0.9$

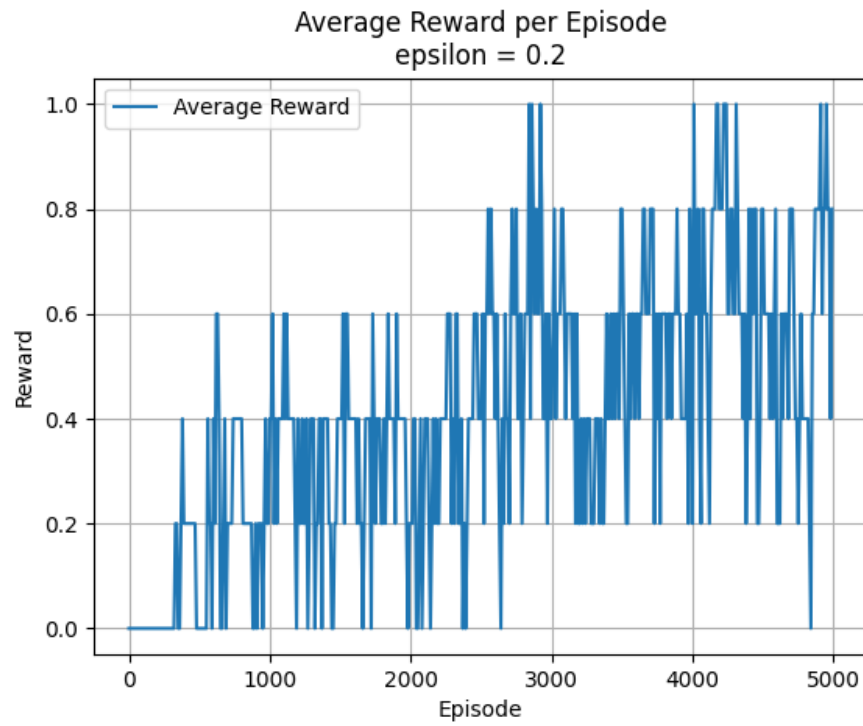


Figure 3: $\epsilon = 0.2$

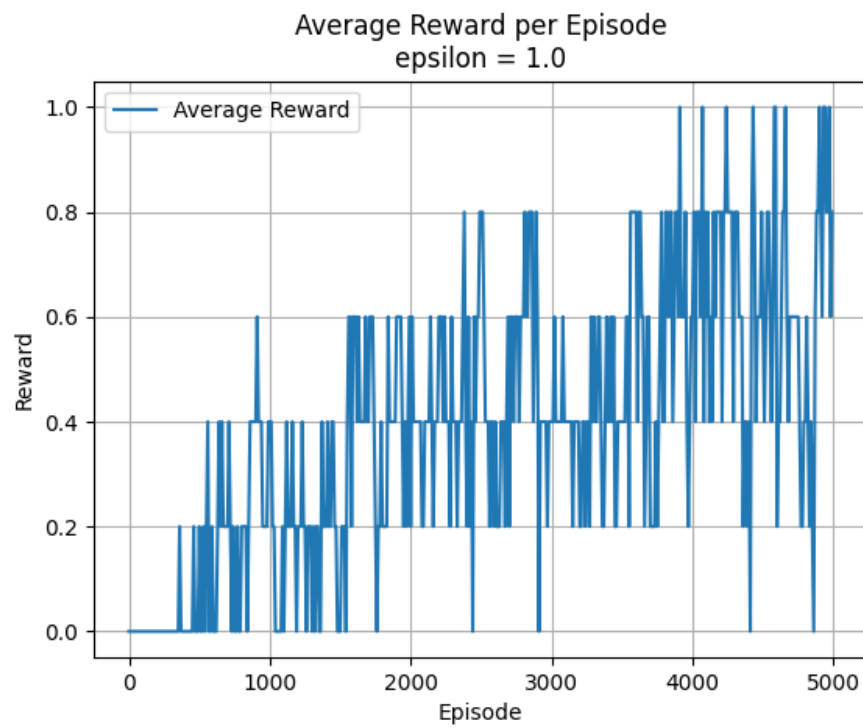


Figure 4: $\epsilon = 1$

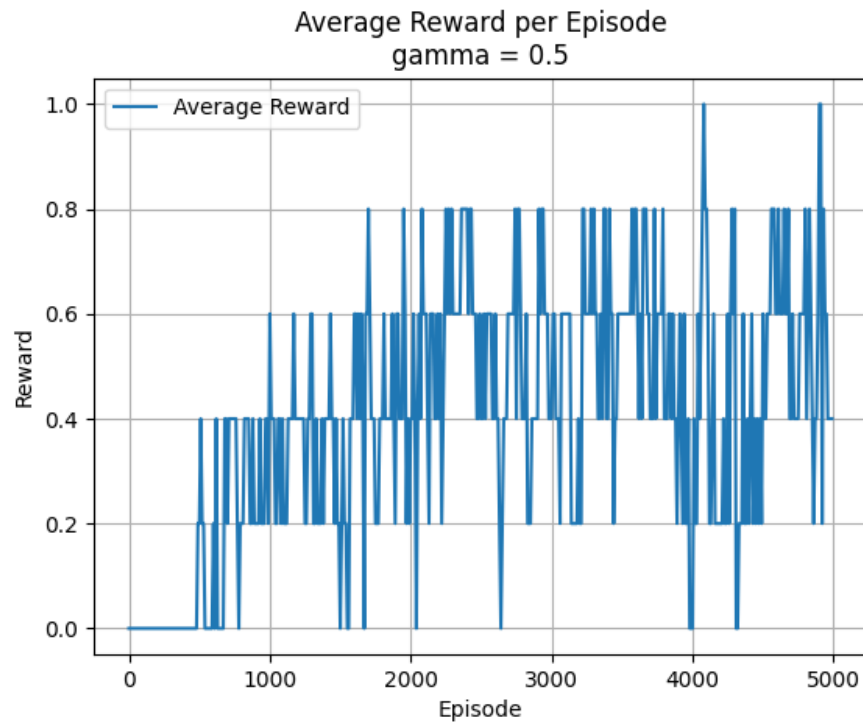


Figure 5: $\gamma = 0.5$

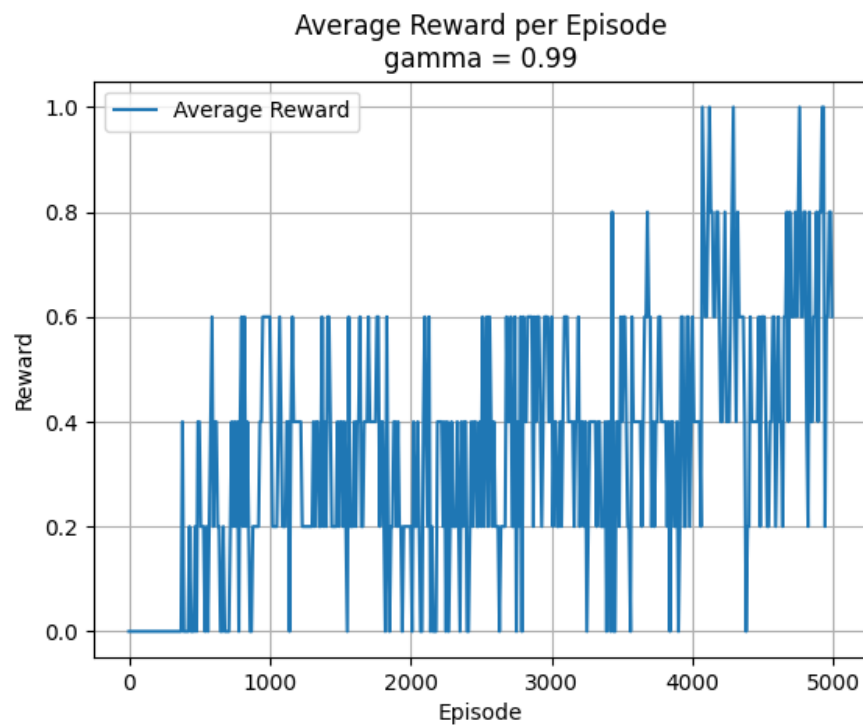


Figure 6: $\gamma = 0.99$

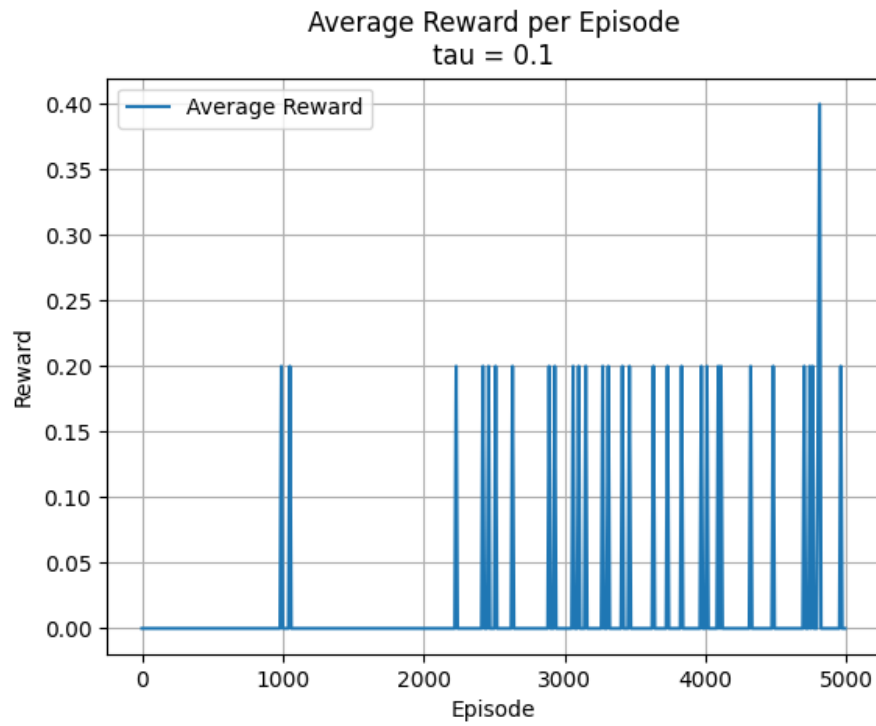


Figure 7: $\tau = 0.1$

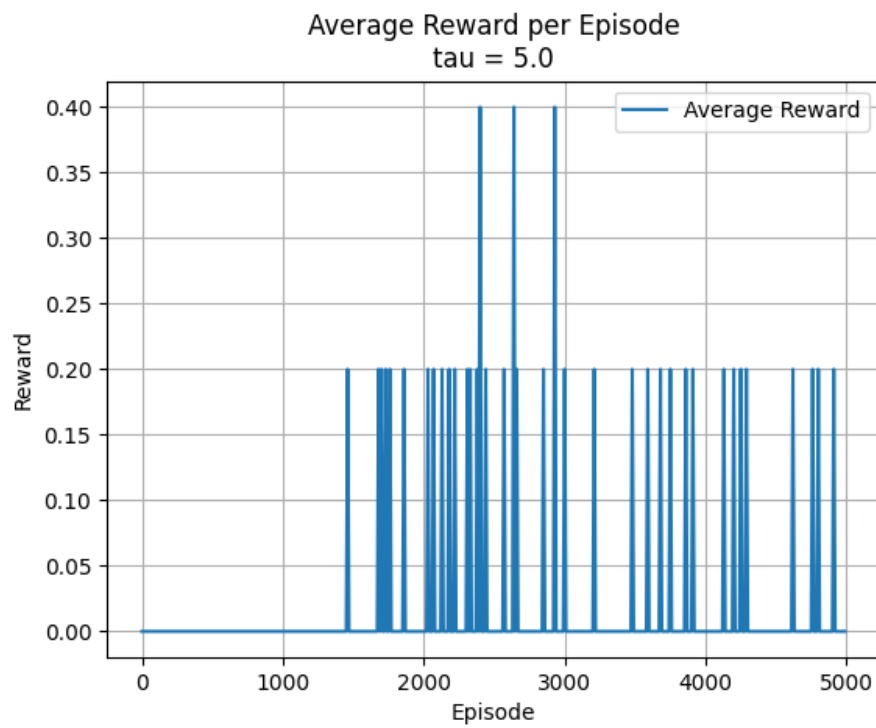


Figure 8: $\tau = 5.0$