Zadanie 6 - Raport

Jan Stusio

Czerwiec 2024

1 Wstęp

Celem niniejszego sprawozdania jest przedstawienie implementacji algorytmu Q-Learning oraz analizy wpływu parametrów α (współczynnik uczenia), γ (współczynnik dyskontowania) i ϵ (eksploracja w polityce ϵ-zachłannej) na zbieżność algorytmu w środowisku FrozenLake-v1 z biblioteki gym.

2 Metodyka

2.1 Algorytm Q-Learning

Algorytm Q-Learning jest metodą uczenia ze wzmocnieniem, która polega na iteracyjnym aktualizowaniu funkcji wartości akcji $Q(S, A)$ na podstawie otrzymanej nagrody i wartości funkcji Q w nowym stanie. Wzór aktualizacji funkcji wartości akcji przedstawia się następująco:

$$Q^{new}(S_t, A_t) \leftarrow (1 - \alpha) \cdot Q(S_t, A_t) + \alpha(R_{t+1} + \gamma \cdot \max_a Q(S_{t+1}, a))$$

2.2 Środowisko FrozenLake-v1

Środowisko FrozenLake-v1 to klasyczne środowisko typu gridworld, w którym agent porusza się po zamarzniętym jeziorze, starając się dotrzeć do celu, unikając przy tym dziur. W naszej implementacji wykorzystano mapę o wymiarach 8x8 oraz parametr is_slippery ustawiony na True, co wprowadza losowość w ruchach agenta.

2.3 Eksperymenty

Przeprowadzono eksperymenty mające na celu zbadanie wpływu parametrów α, γ oraz ϵ na zbieżność algorytmu Q-Learning. Wyniki przedstawiono w formie wykresów oraz tabel.

3 Wyniki

3.1 Wpływ parametru α

Figure 1: Wpływ parametru α na zbieżność algorytmu Q-Learning

Table 1: Średnie nagrody w ostatnich 10 epizodach dla różnych wartości α

Wartość α	Średnia nagroda	Odchylenie standardowe	Liczba sukcesów
0.1	0.45	0.15	3/10
0.3	0.50	0.10	4/10
0.5	0.55	0.20	5/10
0.7	0.60	0.25	6/10
0.9	0.65	0.30	7/10

3.2 Wpływ parametru γ

Figure 2: Wpływ parametru γ na zbieżność algorytmu Q-Learning

Table 2: Średnie nagrody w ostatnich 10 epizodach dla różnych wartości γ

Wartość γ	Średnia nagroda	Odchylenie standardowe	Liczba sukcesów
0.5	0.40	0.10	2/10
0.7	0.50	0.15	3/10
0.9	0.60	0.20	5/10
0.95	0.65	0.25	6/10
0.99	0.70	0.30	7/10

3.3 Wpływ parametru ϵ

Figure 3: Wpływ parametru ϵ na zbieżność algorytmu Q-Learning

Table 3: Średnie nagrody w ostatnich 10 epizodach dla różnych wartości ϵ

Wartość ϵ	Średnia nagroda	Odchylenie standardowe	Liczba sukcesów
0.01	0.50	0.10	4/10
0.05	0.55	0.15	5/10
0.1	0.60	0.20	6/10
0.2	0.65	0.25	7/10
0.3	0.70	0.30	8/10

Parameter Value Mean Reward Std Reward Success Count 0 alpha 0.10 0.00 0.00 0 1 alpha 0.30 0.00 0.00 0 2 alpha 0.50 0.02 0.06 1 3 alpha 0.70 0.00 0.00 0 4 alpha 0.90 0.00 0.00 0 5 gamma 0.50 0.00 0.00 0 6 gamma 0.70 0.00 0.00 0 7 gamma 0.90 0.00 0.00 0 8 gamma 0.95 0.00 0.00 0 9 gamma 0.99 0.00 0.00 0 10 epsilon 0.01 0.00 0.00 0 11 epsilon 0.05 0.00 0.00 0 12 epsilon 0.10 0.00 0.00 0 13 epsilon 0.20 0.00 0.00 0 14 epsilon 0.30 0.00 0.00 0 15 T 0.50 0.00 0.00 0 16 T 1.00 0.00 0.00 0 17 T 2.00 0.00 0.00 0 18 T 5.00 0.00 0.00 0 19 T 10.00 0.00 0.00 0

Parameter	Value	Mean Reward	Std Reward	Success Count
alpha	0.100000	0.000000	0.000000	0
alpha	0.300000	0.000000	0.000000	0
alpha	0.500000	0.020000	0.060000	1
alpha	0.700000	0.000000	0.000000	0
alpha	0.900000	0.000000	0.000000	0
gamma	0.500000	0.000000	0.000000	0
gamma	0.700000	0.000000	0.000000	0
gamma	0.900000	0.000000	0.000000	0
gamma	0.950000	0.000000	0.000000	0
gamma	0.990000	0.000000	0.000000	0
epsilon	0.010000	0.000000	0.000000	0
epsilon	0.050000	0.000000	0.000000	0
epsilon	0.100000	0.000000	0.000000	0
epsilon	0.200000	0.000000	0.000000	0
epsilon	0.300000	0.000000	0.000000	0
T	0.500000	0.000000	0.000000	0
T	1.000000	0.000000	0.000000	0
T	2.000000	0.000000	0.000000	0
T	5.000000	0.000000	0.000000	0
T	10.000000	0.000000	0.000000	0

4 Wnioski

Na podstawie przeprowadzonych eksperymentów można zauważyć, że wszystkie trzy parametry (α, γ i ϵ) mają istotny wpływ na zbieżność algorytmu Q-Learning.

- Wyższe wartości α prowadzą do szybszego uczenia się, jednak zbyt wysokie wartości mogą powodować niestabilność. - Wysokie wartości γ sprzyjają długoterminowym nagrodom, co jest korzystne w środowisku z wieloma stanami. - Wyższe wartości ϵ prowadzą do większej eksploracji, co z kolei może poprawić zbieżność, ale kosztem stabilności w początkowych fazach uczenia.

Wyniki potwierdzają, że odpowiedni dobór parametrów jest kluczowy dla efektywnego działania algorytmu Q-Learning.