WSI-24L-G104

Adam Kwaśnik 22.05.2024

Zadanie 6.

Cel zadania

Celem zadania jest zaimplementowanie przez Państwa algorytmu z podstawami uczenia przez wzmacnianie (Reinforcement learning) w rozwiązaniu problemu Frozen Lake.

Frozen Lake to środowisko dostępne w bibliotece gym – link tutaj. Przykładowe uruchomienie środowiska przedstawione zostało tutaj. Frozen Lake implementuje agenta poruszającego się po lodowej planszy w kierunku celu i unikającego dziurawych pól. Agent może poruszać się w czterech kierunkach w poziomie.

Plansza składa się z czterech typów pól:

Start (S) - pole startowe,
Frozen (F) - bezpieczne pole, po którym agent sie porusza,
Hole (H) - dziura, wejście na to pole kończy grę,
G (Goal) - cel, do którego dąży agent.

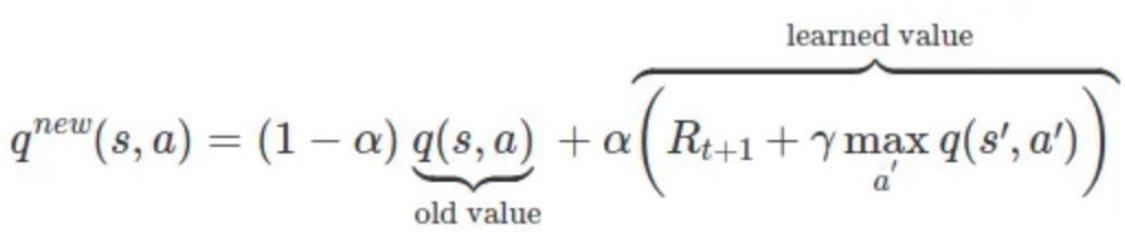
Po uruchomieniu środowiska frozen lake muszą Państwo zaimplementować algorytm Q-learning w celu nauki agenta oczekiwanego sposobu działania.

Algorytm Q-learning powinien obejmować:

- Stworzenie Q - tabeli przechowującej wartość oczekiwanej nagrody dla każdej pary stan-akcja.
- Decyzje agenta - Agent wykorzystuje Q-tabelę do podejmowania decyzji o tym, którą akcję wybrać w danym stanie. W tym celu stosuje strategię epsilon-greedy, która balansuje pomiędzy eksploracją (wybieranie losowych akcji) a eksploatacją (wybieranie najlepszych akcji według Q-tabeli).
- Uczenie przez wzmacnianie - Proces uczenia polega na aktualizacji wartości Q w Q-tabeli na podstawie doświadczeń agenta. Po wykonaniu akcji i otrzymaniu nagrody, agent aktualizuje wartość Q dla danej pary stan-akcja zgodnie z równaniem aktualizacji Q-learningu.

Krok uczenia można opisać jako:

Gdzie:



(https://towardsdatascience.com/q-learning-algorithm-from-explanation-to-implementation-cdbeda2ea187)

- q^{new}(s,a) to nowa wartość funkcji Q po akcji a w stanie *s*,
- α to współczynnik uczenia.
- R_{t+1} to otrzymana nagroda.
- γ to współczynnik dyskontowania określający znaczenie przyszłych nagród w porównaniu z natychmiastowymi,
- $\max_{a'} q(s',a')$ to maksymalna wartość Q spośród wszystkich możliwych do podięcia akcji a' w nowym stanie s'

Sprawozdanie

W sprawozdaniu należy umieścić: opis działania algorytmu oraz zbadać wpływ parametrów: uczenia, dyskontowania i eksploracji na zbieżność algorytmu – momentu gdy algorytm osiąga stabilny stan i dalsze uczenie nie przynosi znaczących zmian.
W sprawozdaniu należy umieścić wykres sumy wszystkich nagród od kroku uczenia.
Należy przedstawić wyniki dla 3 uruchomień algorytmu z różną wartością random seed.