Sprawozdanie z projektu
CliffWalking przy użyciu Q-learningu

Michał Burda
Kamil Poniewierski

Data oddania: 8 maja 2025

1 Wstęp

Celem projektu było zaimplementowanie algorytmu uczenia ze wzmocnieniem w środowisku **CliffWalking-v0** z biblioteki **Gymnasium**. Zadanie polegało na nauczeniu agenta skutecznego poruszania się po planszy, unikając pól klifu, które powodują wysoką karę punktową.

2 Implementacja

W pierwszym kroku zaimportowano potrzebne biblioteki: `gymnasium`, `numpy` oraz `matplotlib`. Następnie utworzono środowisko przy użyciu funkcji:

```
env = gym.make("CliffWalking-v0")
```

Główne parametry algorytmu Q-learningu ustawiono następująco:

- Współczynnik uczenia (α) = 0.1
- Współczynnik dyskontowy (γ) = 0.9
- Początkowe epsilon = 1.0, minimalne epsilon = 0.01, współczynnik zmniejszania epsilon = 0.995

Q-tablica została zainicjalizowana jako macierz zer o wymiarach odpowiadających liczbie stanów i akcji w środowisku.

Uczenie odbywało się w pętli obejmującej 500 epizodów. W każdym epizodzie agent:

- Wybierał akcję metodą **epsilon-greedy**.
- Otrzymywał nagrodę i aktualizował swoją pozycję.
- Aktualizował wartość Q według formuły Q-learningu.
- Zmniejszał wartość epsilon po każdym epizodzie.

3 Wykorzystany algorytm: Q-learning

Q-learning jest metodą model-free uczenia ze wzmocnieniem, w której agent uczy się optymalnej polityki wyboru akcji na podstawie nagród.

Aktualizacja Q-wartości odbywa się według wzoru:

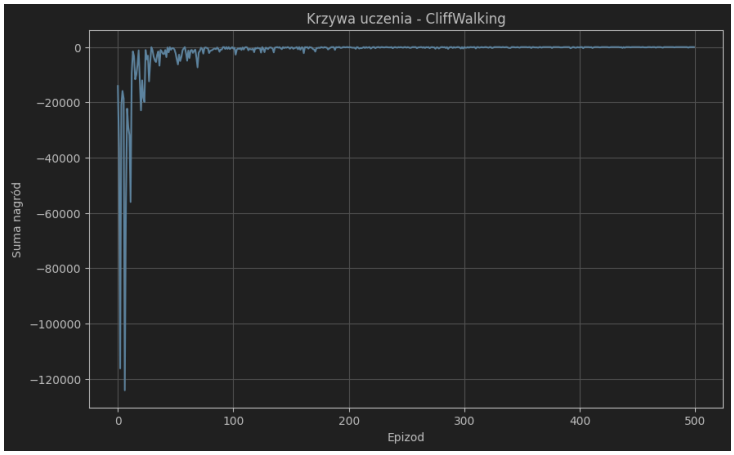
$$Q(s,a) \leftarrow Q(s,a) + \alpha \left(r + \gamma \max_{a'} Q(s',a') - Q(s,a) \right)$$

gdzie:

- s - obecny stan,
- a - podjęta akcja,
- r - otrzymana nagroda,
- s' - nowy stan po wykonaniu akcji.

4 Eksperymenty i wyniki

W trakcie eksperymentów agent uczył się skutecznie poruszać po planszy, omijając klify. Proces treningu był monitorowany poprzez sumowanie nagród w każdym epizodzie.



Rysunek 1: Enter Caption

Rysunek 2: Krzywa uczenia agenta w środowisku CliffWalking

Początkowo agent często wpadał w klif, otrzymując wysokie kary. Jednak w miarę zmniejszania wartości epsilon i dalszego treningu suma nagród zaczęła rosnąć, co świadczyło o poprawie strategii.

5 Wnioski

Zaimplementowany algorytm Q-learningu skutecznie nauczył agenta omijania przeszkód i dotarcia do celu w środowisku CliffWalking. Eksperyment pokazał znaczenie równowagi między eksploracją a eksploatacją oraz dobrania odpowiednich parametrów algorytmu.

Projekt umożliwił praktyczne zrozumienie mechanizmu uczenia ze wzmocnieniem oraz działania algorytmu Q-learning w środowisku o dyskretnej przestrzeni stanów.