Sprawozdanie

WSI – Ćwiczenie 6: Uczenie się ze wzmocnieniem

Yan Korzun

12 stycznia 2026

1 Cel ćwiczenia

Celem ćwiczenia była implementacja algorytmu **Q-learning** z uczeniem epizodycznym oraz strategią wyboru akcji typu ε-zachłanna. Zaimplementowany algorytm został zastosowany do wytrenowania agenta w środowisku *MiniGrid Four Rooms*. Dodatkowym celem było zbadanie wpływu parametrów algorytmu na jakość oraz szybkość uczenia się agenta.

2 Opis środowiska

Eksperymenty przeprowadzono w środowisku `MiniGrid-FourRooms-v0`, które składa się z czterech pomieszczeń oddzielonych ścianami z przejściami. Zadaniem agenta jest dotarcie do pola celu w jak najmniejszej liczbie kroków. W celu uproszczenia problemu przyjęto, że pozycja agenta oraz pozycja celu są stałe, a maksymalna liczba kroków w epizodzie jest ograniczona.

3 Zastosowany algorytm

Algorytm Q-learning aktualizuje wartości funkcji jakości akcji zgodnie ze wzorem:

$$Q(s,a) \leftarrow Q(s,a) + \beta \left(r + \gamma \max_{a'} Q(s',a') - Q(s,a) \right) \tag{1}$$

gdzie s oznacza aktualny stan, a wybraną akcję, r otrzymaną nagrodę, s' kolejny stan, β współczynnik uczenia, a γ współczynnik dyskontowania.

Do wyboru akcji zastosowano strategię ε-zachłanną, polegającą na losowym wyborze akcji z prawdopodobieństwem ε lub wyborze akcji o najwyższej wartości Q z prawdopodobieństwem $1 - \varepsilon$.

4 Plan eksperymentów

Przeprowadzono serię eksperymentów dla różnych zestawów parametrów algorytmu, w szczególności:

- współczynnika dyskontowania γ,
- współczynnika uczenia β,

- parametru eksploracji ε.

Dla każdej konfiguracji agent był trenowany przez ustaloną liczbę epizodów. Jako miary jakości uczenia wykorzystano średnią liczbę kroków potrzebnych do osiągnięcia celu oraz średnią sumę nagród w epizodzie.

5 Wyniki eksperymentów

Wyniki przeprowadzonych eksperymentów przedstawiono w tabeli 1.

Tabela 1: Wyniki uczenia algorytmu Q-learning dla różnych parametrów

γ	β	ε	Średnia nagroda	Epizody
0.99	0.2	0.4	0.301	500
0.99	0.2	0.4	0.948	1000
0.99	0.2	0.4	0.958	1500
0.99	0.4	0.4	0.000	500
0.99	0.4	0.4	0.000	1000
0.99	0.4	0.4	0.000	1500
0.99	0.4	0.8	0.315	500
0.99	0.4	0.8	0.905	1000
0.99	0.4	0.8	0.95	1500
0.98	0.4	0.8	0.316	500
0.98	0.4	0.8	0.918	1000
0.98	0.4	0.8	0.962	1500
0.98	0.2	0.4	0.003	500
0.98	0.2	0.4	0.000	1000
0.98	0.2	0.4	0.000	1500

6 Analiza wyników

Przeprowadzone eksperymenty wykazały, że skuteczność algorytmu Q-learning w środowisku MiniGrid-FourRooms silnie zależy od doboru parametrów eksploracji. Wartość parametru powinna być duża żeby agent dotarł do celu. Natomiast zbyt mała uniemożliwia odkrycie celu. Współczynnik musi być wystarczająco duży, opdowiedni do liczby kroków. Najlepsze wyniki uzyskano dla wysokiego współczynnika dyskontowania (= 0.99) oraz umiarkowanej szybkości uczenia ($\beta = 0.2$).

7 Wnioski

W ramach ćwiczenia poprawnie zaimplementowano algorytm Q-learning oraz zastosowano go do rozwiązania problemu Four Rooms. Agent nauczył się skutecznie docierać do celu,

a przeprowadzone eksperymenty potwierdziły istotny wpływ parametrów algorytmu na jakość uczenia się.