Autorka: Patrycja Wysocka
Przedmiot: Wstęp do Sztucznej Inteligencji
Nr albumu: 306 022

Ćwiczenie 6 – Uczenie się ze wzmocnieniem

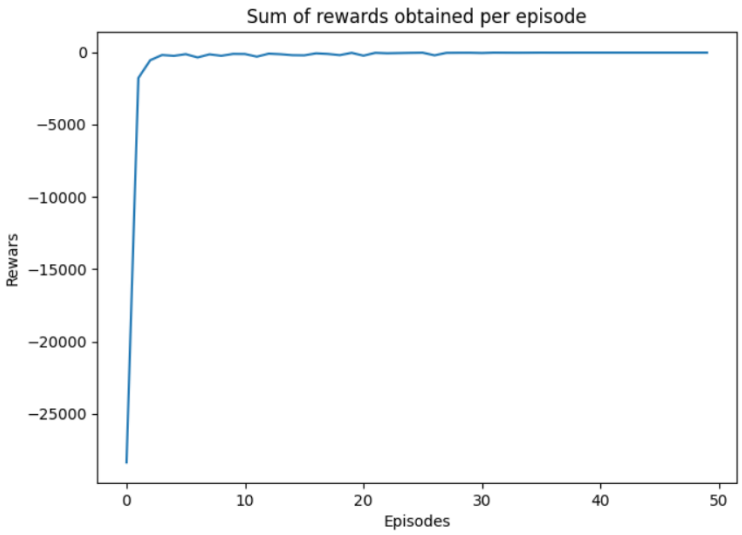
1) Treść zadania

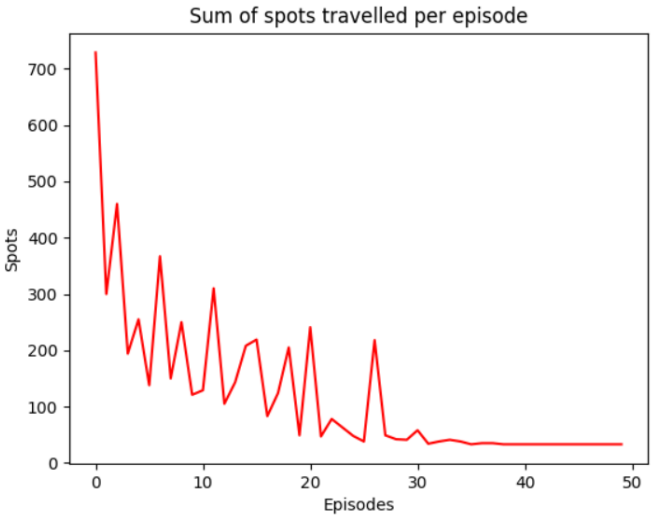
Zaimplementować algorytm **Q-Learning**. Zebrać i przedstawić na wykresie liczbę wykonanych kroków i naliczoną karę/nagrodę w kolejnych epokach. Problem do rozwiązania to znalezienie drogi z punktu 'S' do punktu 'F' w "labiryncie" / świecie z przeszkodami. Rezultatem działania algorytmu powinna być ścieżka w postaci: (1,1)->(0,1)->...->(2,3) oraz ww. wykres. Przykładowe mapy powinny być czytane na starcie programu z jakiegoś formatu np. ASCI (gdzie '#'-przeszkoda).

2) Przyjęte założenia

- Algorytm może „przechodzić przez ściany” (oczywiście otrzymuje wtedy odpowiednio wysoką karę)
- Dany epizod kończy się w momencie gdy algorytm dotarł do celu
- W procesie uczenia się algorytm zawsze startuje z tego samego miejsca
- Algorytm planując kolejny krok zawsze podejmuje najlepszą (wg niego) decyzję - brak jakichkolwiek ruchów losowych
- Learning rate = 0.85
- Discount factor (beta) = 0.8
- Testowany labirynt znajduje się w „maze1.txt”
- ‘#’ jest uznawany jako ściana, ‘.’ jako wolne pole

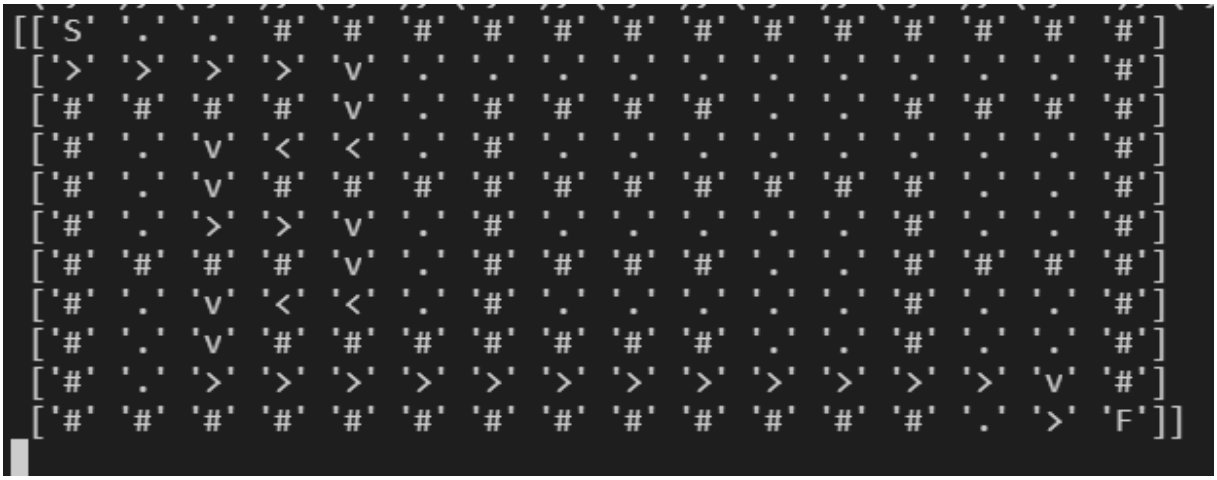
3) Raport z przeprowadzonych eksperymentów





Ścieżka: [(0, 0), (1, 0), (1, 1), (1, 2), (1, 3), (1, 4), (2, 4), (3, 4), (3, 3), (3, 2), (4, 2), (5, 2), (5, 3), (5, 4), (6, 4), (7, 4), (7, 3), (7, 2), (8, 2), (9, 2), (9, 3), (9, 4), (9, 5), (9, 6), (9, 7), (9, 8), (9, 9), (9, 10), (9, 11), (9, 12), (9, 13), (9, 14), (10, 14), (10, 15)]

Graficznie pokazany labirynt:



4. Obserwacje i wnioski

- Na samym początku algorytm wykonuje bardzo dużo kroków oraz otrzymuje ogromną karę – dzieje się tak dlatego, że dopiero poznaje labirynt i często wpada na ściany
- Różnice pomiędzy otrzymanymi nagrodami są początkowo bardzo duże – dzieje się tak dlatego, że algorytm przy swoich pierwszych przejściach przez labirynt trafia na najwięcej ścian. Odpowiednia duża funkcja kary zapewnia jednak, że przy każdym kolejnym epizodzie algorytm je omija – stąd tak gwałtowne zwiększenie liczby nagród.
- Mimo ustabilizowania się wartości nagród, wciąż trwają wahania liczby wykonanych kroków – obrazuje to etap, na którym algorytm nauczył się już nie trafiać na ściany, natomiast nie opracował jeszcze najkrótszej ścieżki i chodzi ‘naokoło’, eksplorując wolne pola.