

Politechnika Warszawska

Wprowadzenie do sztucznej inteligencji

Ćwiczenie 6

Zaimplementować algorytm Q-Learning. Problem do rozwiązania to znalezienie drogi z punktu 'S' do punktu 'F' w "labiryncie" / świecie z przeszkodami.

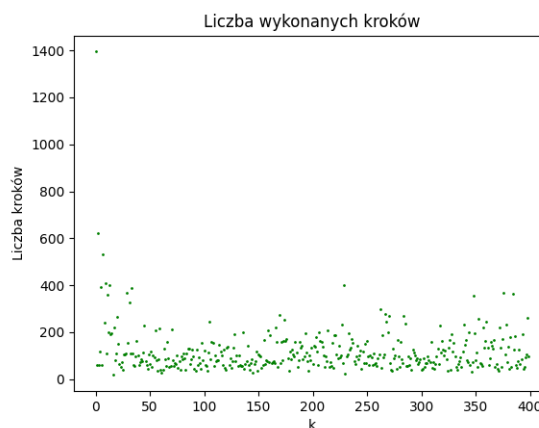
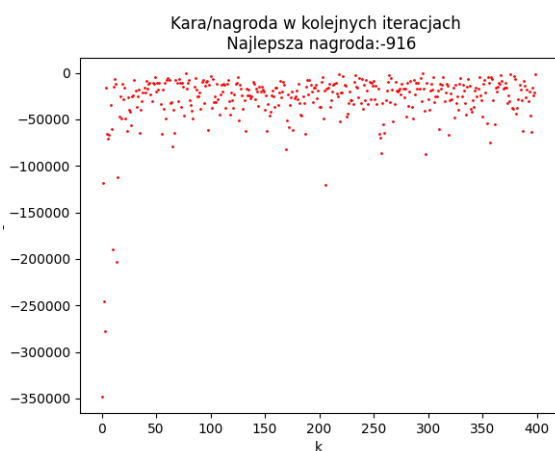
Przemysław Krasnodębski

Link do repozytorium: [WSI-21Z/Cwiczenie 6 at master · p-krasnodebski/WSI-21Z \(github.com\)](https://github.com/p-krasnodebski/WSI-21Z)

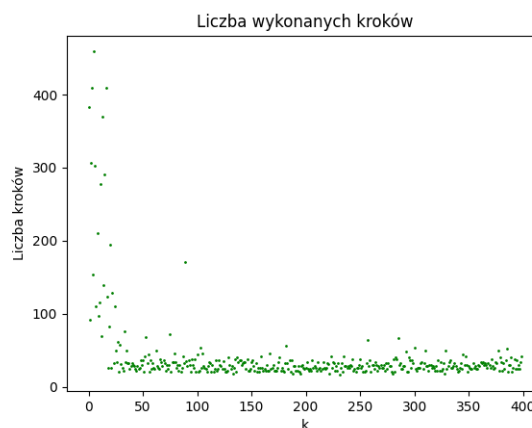
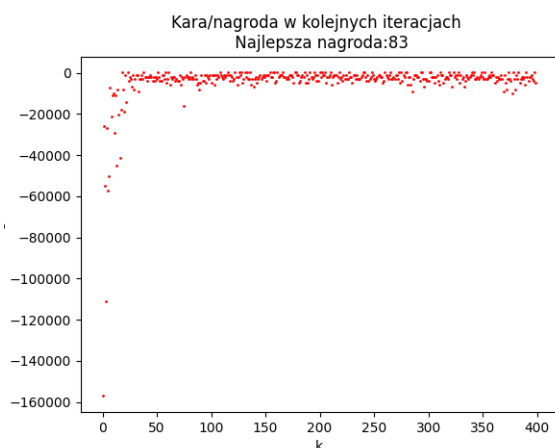
W ramach realizacji zadania zaimplementowano algorytm wskazujący drogę wyjścia z labiryntu, który wczytany zostaje z pliku *maze.txt*. Algorytm może realizować przejście z dowolnie wyznaczonych przez gracza (legalnych) punktów lub z losowych wyznaczonych przez algorytm. Zdefiniowano następujące kary za przebywanie w labiryncie: -1 za pobyt w wolnej przestrzeni w labiryncie, -1000 za pobyt w miejscu ściany algorytmu. Algorytm może wyznaczać ścieżki przez ściany, lecz wiąże się to ze stosunkowo dużą karą. Nagroda za znalezienie wyjścia wynosi 100.

Na poniższych wykresach przedstawiono wyniki działania algorytmu dla różnego parametru epsilon:

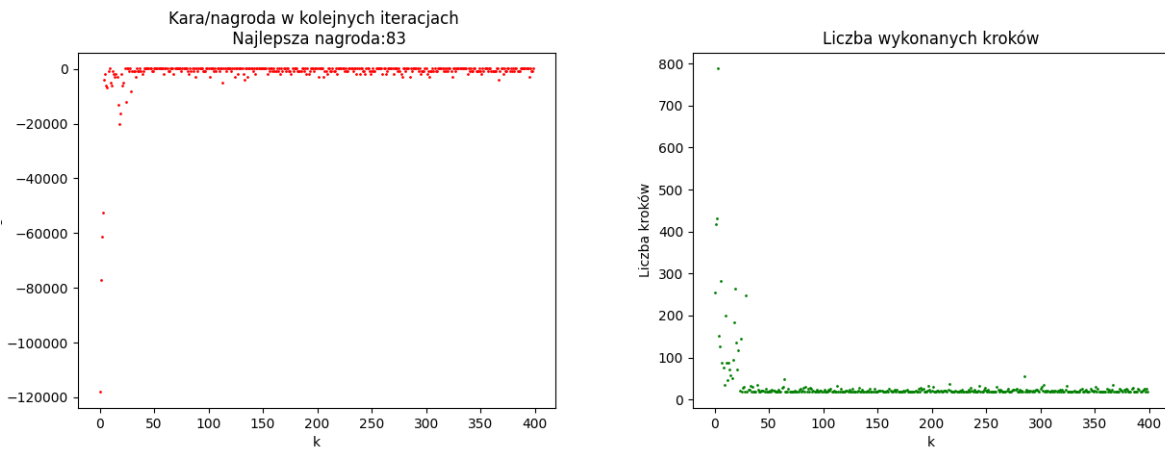
Epsilon = 0.3



Epsilon = 0.7



Epsilon = 0.9



Parametr epsilon ma dość istotny wpływ na działanie algorytmu. Definiuje on prawdopodobieństwo z jakim wybierane są przyszłe akcje. Dla prawdopodobieństwa poniżej progu akcje wybierane są zgodnie z macierzą Q , w przeciwnym przypadku akcja jest losowa. Ma on główny wpływ na eksplorację oraz eksploatację. Dla mniejszych wartości epsilon nacisk kładziony jest na eksplorację, czyli poznawanie skutków różnych akcji (niekoniecznie najlepszych). Zbyt mała wartość parametru powoduje zbyt dużą chaotyczność algorytmu i marne efekty. Natomiast duża wartość epsilon zwiększa nacisk algorytmu na eksploatację. Algorytm wybiera głównie akcje z największymi nagrodami. Może to powodować błędne działanie i brak możliwości znalezienia lepszego rozwiązania. Parametr jaki wybrano to 0.7, gdyż dla większości przypadków uzyskiwane efekty są bardzo dobre, a algorytm ma większe szanse na znalezienie optimum globalnego dla bardziej skomplikowanych przypadków.

Pozostałe parametry algorytmu, są unikalne dla każdego problemu i ich wyznaczenie wykonano bez udokumentowania w sprawozdaniu.

Zaimplementowany algorytm skutecznie realizuje proces uczenia się i znajdowanie drogi w labiryncie. Dla testowanych labiryntów istniejące rozwiązanie znajdowane jest już po 50 iteracji. Jako optymalne rozwiązanie wybierana jest najlepsza ścieżka z dostępnych, gdyż ostatnia iteracja mogła być wykonana z dużą częścią losowych ruchów, co oznacza, że najprawdopodobniej nie jest optymalna.