WSI 21Z

Wojciech Gierulski

Laboratorium 6 - QLearning

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.

Założenia wstępne

Przyjęte założenia:

- Do generacji sterowania użyto strategii ε-zachłannej, gdzie $\varepsilon = 0{,}1$
- Przyjęto współczynnik uczenia $\beta = 0{,}2$ – dobór eksperymentalny
- Przyjęto współczynnik dyskontowania $\gamma = 0{,}8$ – dobór eksperymentalny
- Epoka kończy się, gdy agent dojdzie do punktu końcowego labiryntu lub wykona 200 kroków
- Liczba epok – 500 – dobór eksperymentalny
- Testowany labirynt (**S** - start, **F** – finish, **#** - przeszkoda, **.** – wolne pole):

```
S..##########
...#........#
#..#######..#
#..#........#
#..#..#######
#........#..#
####..#..#..#
#.....#.....
#########..F
```

- Nagrody:
  - Dojście do celu: +5
  - Każda chwila spędzona w labiryncie: -0,1
  - Uderzenie w przeszkodę: -0,5
  - Wejście na odwiedzone już pole: -1

Wyniki eksperymentów

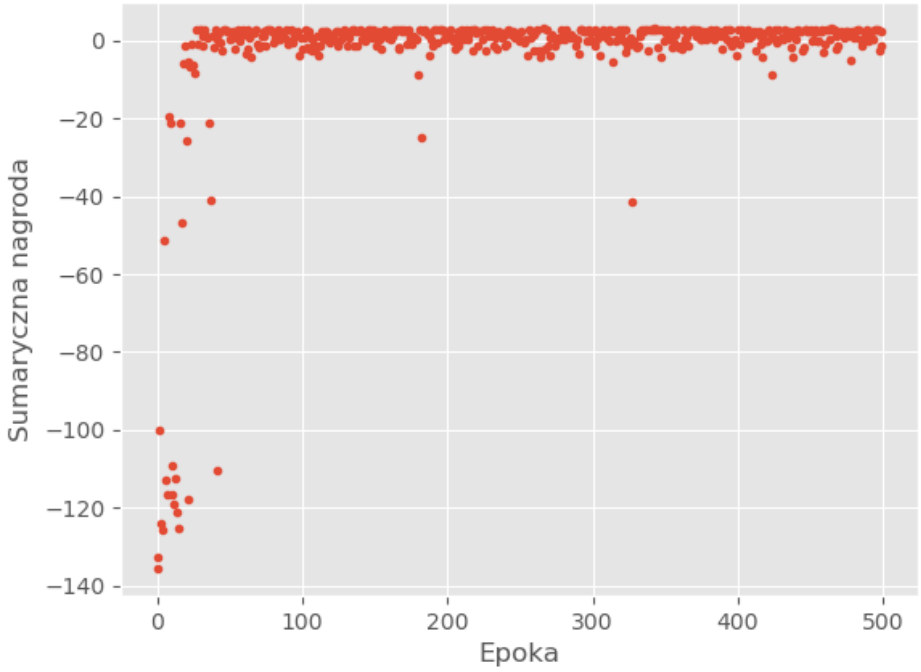
Wynikiem działania programu jest zestaw instrukcji prowadzących z punktu S do F i trasa w notacji (nr rzędu, nr kolumny). Wynik działania programu jest poprawny dla powyższego labiryntu:

[(0, 0), (0, 1), (0, 2), (1, 2), (2, 2), (2, 1), (3, 1), (4, 1), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (5, 7), (6, 7), (7, 7), (7, 8), (7, 9), (7, 10), (7, 11), (7, 12)]

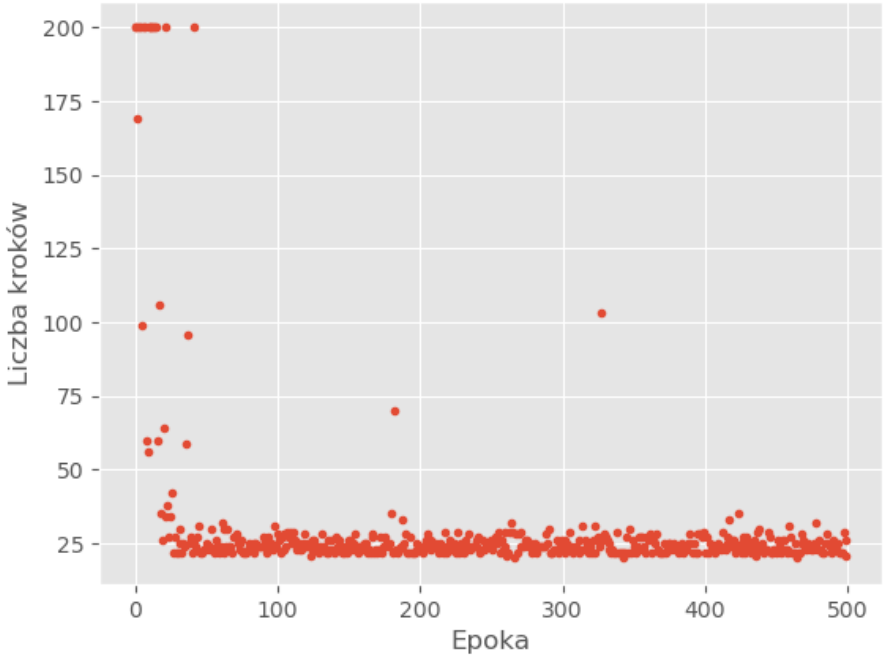
['RIGHT', 'RIGHT', 'DOWN', 'DOWN', 'LEFT', 'DOWN', 'DOWN', 'DOWN', 'RIGHT', 'RIGHT', 'RIGHT', 'RIGHT', 'RIGHT', 'RIGHT', 'DOWN', 'DOWN', 'RIGHT', 'RIGHT', 'RIGHT', 'RIGHT', 'RIGHT', 'DOWN']

Wygenerowano też wykresy sumy nagród i wykonanych kroków w kolejnych epokach. Dane przedstawiono na rysunkach 1 i 2. Zachowanie agenta w środowisku jest intuicyjne – na początku epoki

kończone są z powodu przekroczenia limitu kroków (200) – losowe kroki nie dają dobrych rezultatów, a suma nagród jest niska (około -120). Po kilkudziesięciu epokach agent zaczyna dochodzić do celu kończąc epokę wcześniej (wskazuje na to zmniejszająca się liczba kroków w kolejnych epokach), a suma nagród zwiększa się do około 0.



Rys. 1 – sumaryczna nagroda w kolejnych epokach



Rys. 2- liczba kroków w kolejnych epokach