

Wprowadzenie do sztucznej inteligencji - ćwiczenie 6

Zadanie

Proszę zaimplementować algorytm Q-Learning i użyć go do wyznaczenia polityki decyzyjnej dla problemu FrozenLake8x8-v0 (w wersji domyślnej, czyli z włączonym poślizgiem). W problemie chodzi o to, aby agent przedostał się przez zamrożone jezioro z pozycji 'S' do pozycji 'G' unikając punktów 'H'. Symulator dla tego problemu można pobrać z podanej strony lub napisać własny o takiej samej funkcjonalności. Oprócz zbadania domyślnego sposobu nagradzania (1 za dojście do celu, 0 w przeciwnym przypadku) proszę zaproponować własny system nagród i kar, po czym porównać osiągnięte wyniki z wynikami systemu domyślnego.

Za wynik uznajemy procent dojść do celu w 1000 prób. W każdej próbie można wykonać maksymalnie 200 akcji.

Rozwiązanie

Przy implementacji zadania wykorzystano symulator problemu z podanej strony w wersji domyślnej. Zaimplementowano algorytm QLearning z wyborem następnego ruchu za pomocą metody epsilon greedy. Zgodnie z zadaniem porównano trzy funkcję nagrody:

default_reward:

- 1 pkt za dojście do celu i 0 w p.p

dont_like_holes_reward:

- 1 pkt za dojście do celu, - 1 pkt za wpadnięcie w dziurę, 0 w p.p

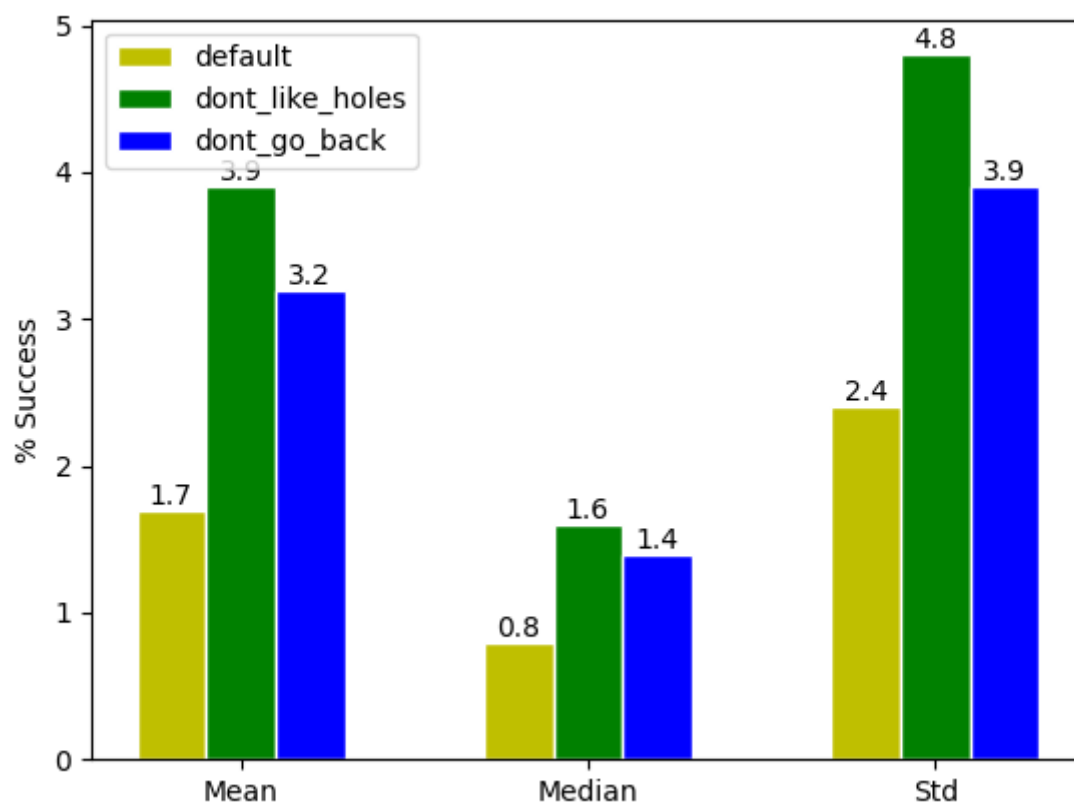
dont_go_back_reward:

- 1 pkt za dojście do celu, -1 pkt za wpadnięcie w dziurę i powrót na start, 0 w p.p

Pozostałe parametry dla których testowano to:

- liczba iteracji: 1000,
- maksymalna liczba ruchów: 200,
- leaning_rate: 0.1,
- discount_rate: 0.1,
- epsilon: 0.01.

Wyniki dla 500 uruchomień każdej z funkcji nagrody prezentują się następująco:



Wnioski

Możemy zauważyć, że najlepiej spisał się algorytm korzystający z nagrody za dojście do celu oraz kary za wpadnięcie w dziurę. Co ciekawe algorytm, który był dodatkowo karany za powrót na start poradził sobie od niego nieco gorzej, chociaż i tak znacznie lepiej niż ten z nagrodą domyslną. Należy też zwrócić uwagę, że są to wyniki uśrednione, z powodu dużej losowości algorytmu. Każdemu z nich zdarzały się epizody, gdy dojścia do celu były równe 0, a najlepszy z nich potrafił czasem osiągać wyniki nawet rzędu 25%.