

# WSI - Q-learning

Adam Sokołowski

Maj 2024

## 1 Cel ćwiczenia

Implementacja algorytmu Q-learning i stworzenie agenta rozwiązującego problem Taxi z pakietu *gynasium*. Następnie zbadanie wpływu parametrów oraz przedstawienie wyników.

## 2 Wyniki eksperymentów

### 2.1 Test wpływu parametru beta (learning rate)

Poprzez regulację beta, można kontrolować tempo uczenia się agenta. Większe beta może pomóc szybciej dostosować się do zmieniającego się środowiska, ale może również prowadzić do większych fluktuacji w wartościach Q. Mniejsze beta może zapewnić stabilniejsze uczenie się, ale może też spowodować wolniejsze dostosowywanie się do nowych informacji.

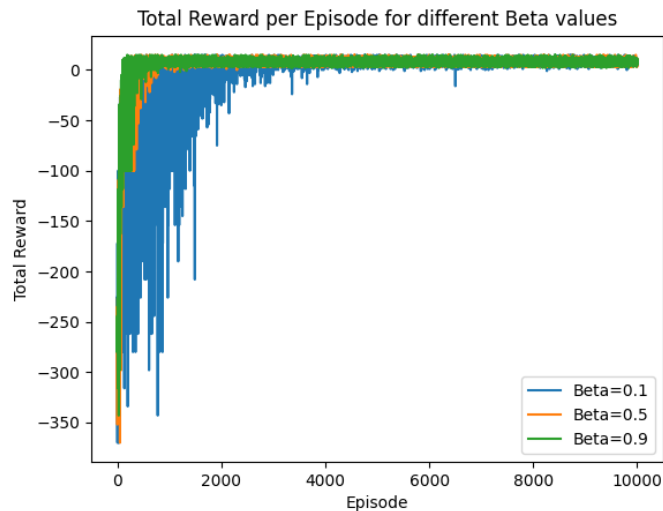


Figure 1: Test beta

Jak widać im większy parametr beta tym szybciej agent się uczy, jednak w tym testie ciężko było pokazać wcześniej wspomniane fluktuacje w wartościach Q. Zatem najbezpieczniejszym wyborem byłaby beta równa 0.5.

## 2.2 Test wpływu parametru gamma (współczynnikiem dyskontującym)

Gamma określa, jak dużą wagę algorytm Q-learning przywiązuje do przyszłych nagród w porównaniu do bieżących nagród. Zbyt niskie gamma może sprawić, że algorytm będzie ignorował przyszłe korzyści. Zbyt wysokie gamma może sprawić, że agent będzie silnie preferował długoterminowe nagrody, co może prowadzić do zbyt dużego odwlekania nagród. W praktyce wartości gamma są zazwyczaj wybierane w przedziale od 0.9 do 0.99, co oznacza, że agent bierze pod uwagę przyszłe nagrody, ale z pewnym stopniem dyskontowania.

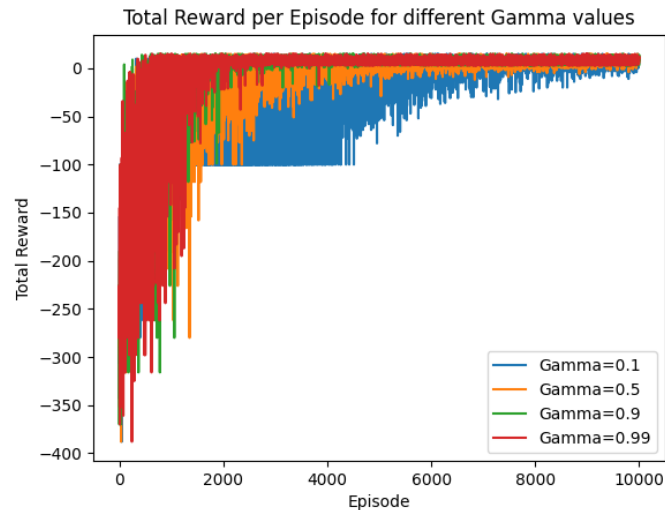


Figure 2: Test gamma

Jak widać najlepsza wartość dla parametru gamma to 0.99.

## 2.3 Test końcowy

Na poniższym rysunku pokazano proces uczenia się agenta dla wcześniej dobranych parametrów. Przyjęto 10000 epizodów i maksymalnie 100 kroków na epizod.

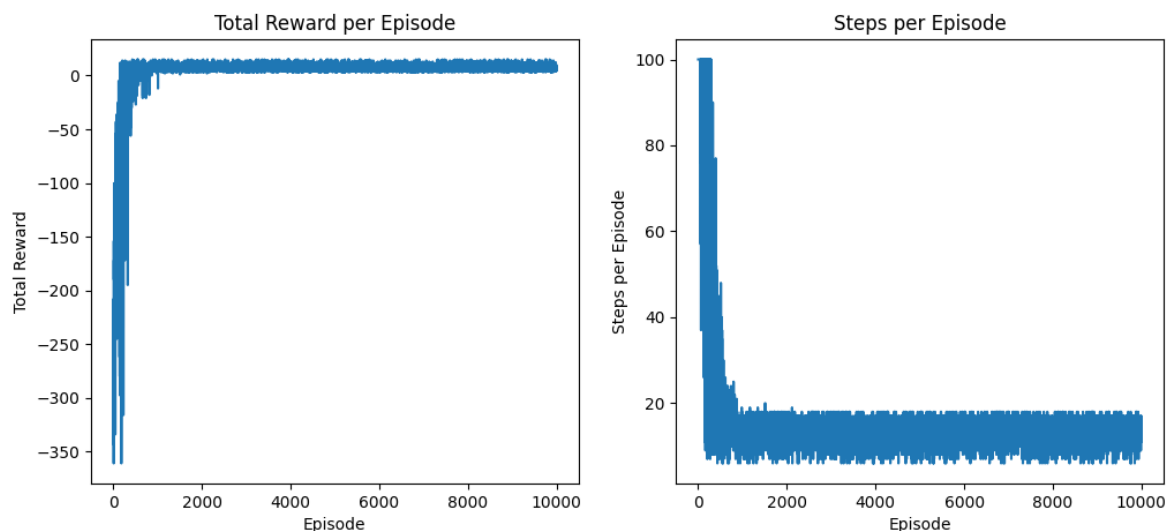


Figure 3: Test gamma