Wydział Elektroniki i Technik Informacyjnych
Politechnika Warszawska

Wprowadzenie do sztucznej inteligencji

Sprawozdanie z ćwiczenia nr 6

Tymon Kobylecki

Warszawa, 2022

Spis treści

1. Wstęp

W niniejszym sprawozdaniu opisane zostało rozwiązanie zadania oraz eksperymenty dotyczące zadania nr 6 polegającego na implementacji algorytmu Q-learning. Miał on za zadanie rozwiązywać problem Taxi z pakietu `gym`, dostępny pod adresem `https://web.archive.org/web/20210125043510/http://gym.openai.com/envs/Taxi-v3/`.

2. Ćwiczenie

2.1. Środowisko - problem Taxi

W dostarczonym środowisku taksówka miała za zadanie przewozić pasażerów między 2 z 4 możliwych punktów umieszczonych wewnątrz labiryntu na zorientowanej mapie o wymiarach 5 na 5 pól. Taksówka w każdym momencie miała do wyboru 6 ruchów:

— 0 - ruch na południe
— 1 - na północ
— 2 - na wschód
— 3 - na zachód
— 4 - pobranie pasażera
— 5 - wysadzenie pasażera

Taksówka ma możliwość wykonywania ruchów nieprawidłowych, np. wjeżdżania w ścianę albo pobieranie pasażerów tam, gdzie ich nie ma.

2.2. Eksperymenty

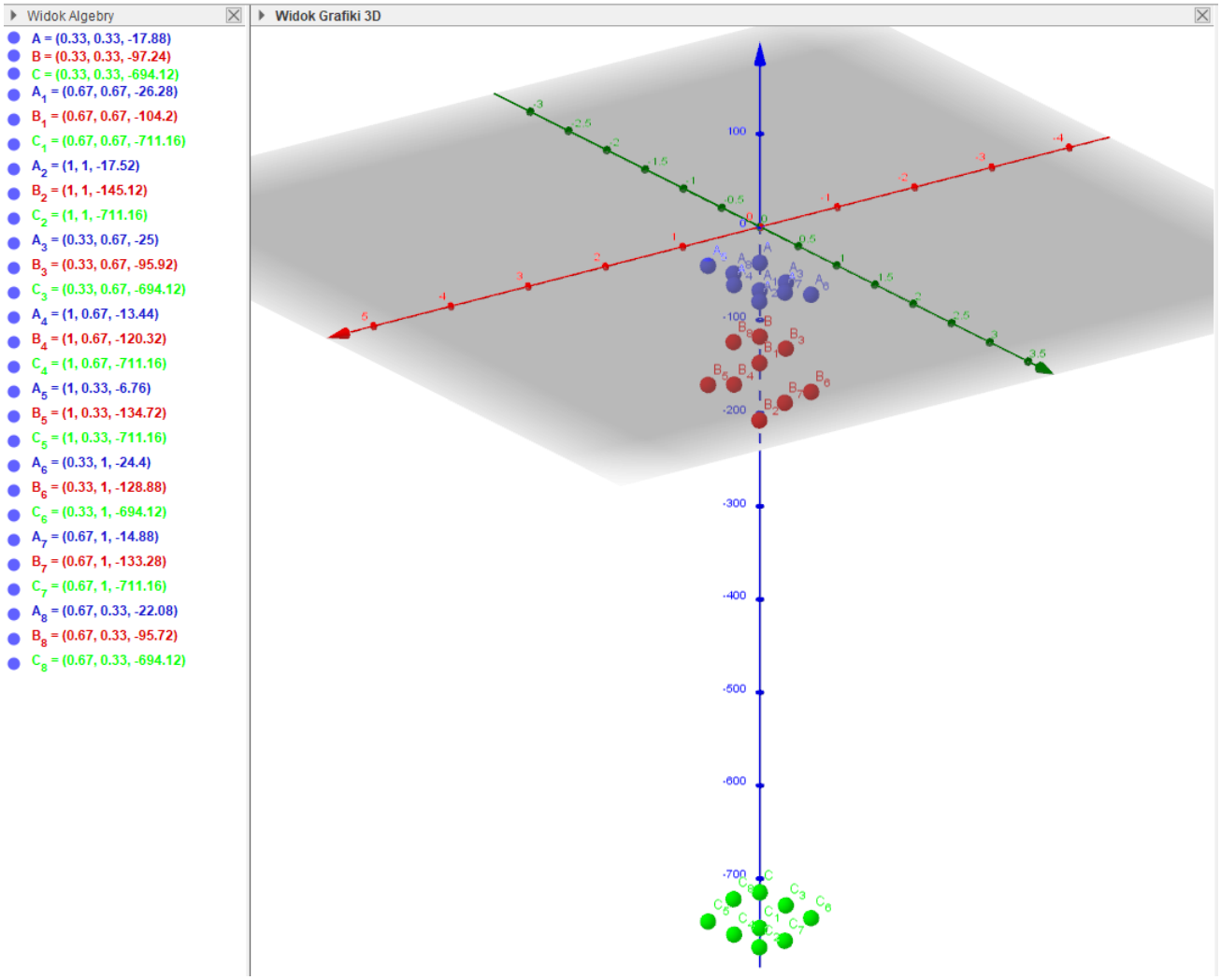
Eksperymentom poddane zostały parametry learning rate, γ oraz prawdopodobieństwa wybrania losowej polityki. Z uwagi na skończoność czasu ludzkiego życia zbiór wartości został ograniczony do $[0{,}33, 0{,}67, 1]$. Liczba pokoleń była w każdym eksperymencie jednakowa i wynosiła 500. Rezultatem eksperymentów były średnie z 25 uruchomień algorytmu dla danych wartości. Dla uczciwości eksperymentów scenariusze środowiska zostały powtórzone dla wszystkich eksperymentów, czyli było 25 różnych środowisk, powtórzonych dla wszystkich parametrów.

2.3. Wyniki

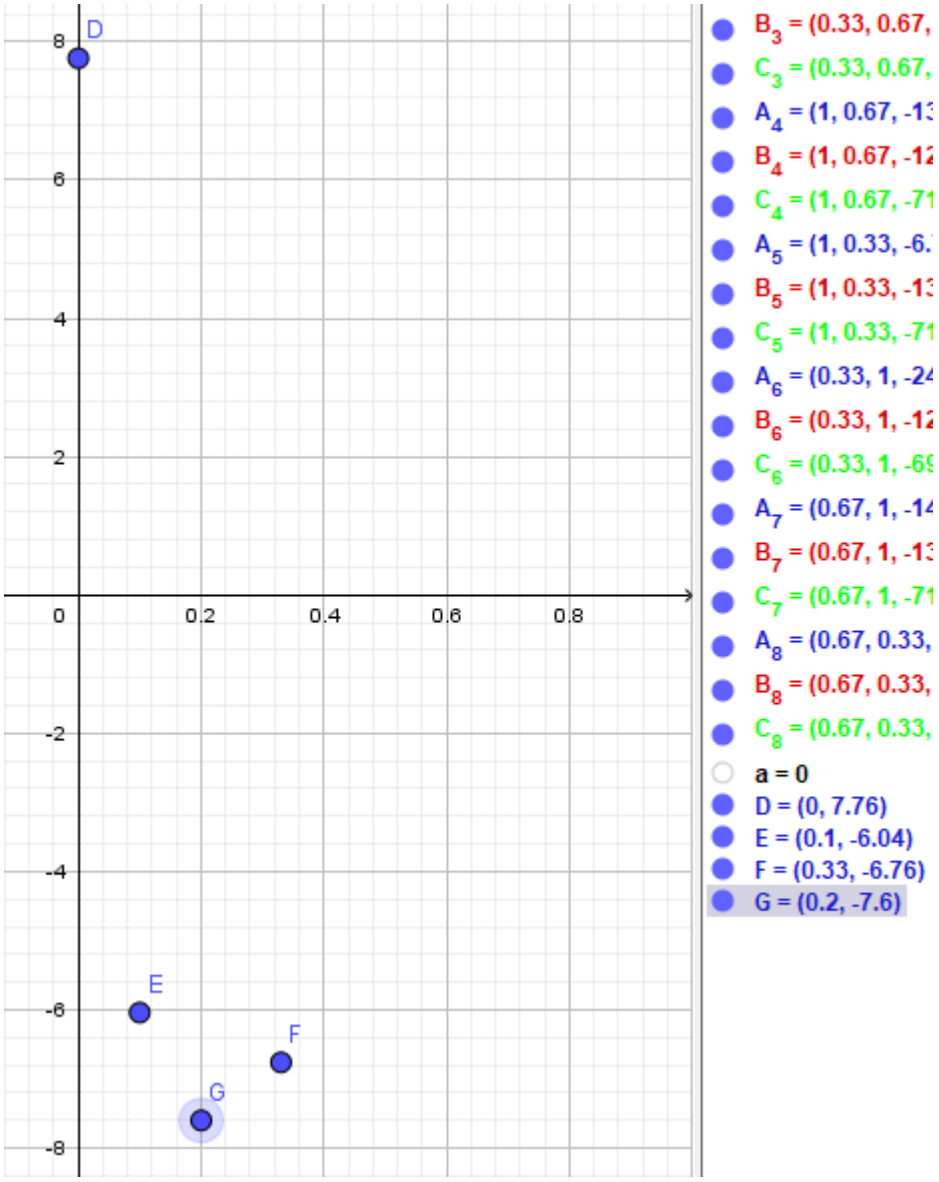
Na trójwymiarowym wykresie 2.1 zostały przedstawione wyniki 27 eksperymentów. Na osi x (czerwonej) znajduje się parametr learning rate, oś y (zielona) symbolizuje parametr γ, czerwona - średni wynik z 25 uruchomień algorytmu, zaś prawdopodobieństwo wyboru losowej polityki zostało, z uwagi na trudność wizualizacji 4-wymiarowego wykresu, przedstawione za pomocą kolorów - niebieski to 0,33, czerwony to 0,67, zaś zielony oznacza 1.

2.4. Analiza wyników

Łatwo zauważyć, że wraz z obniżonym prawdopodobieństwem wyraźnie wzrasta średnia wyników. Pozostałe parametry, w porównaniu do prawdopodobieństwa, minimalnie wpływają na rezultat, jednak widocznym maksimum jest punkt symbolizujący learning rate równe 1, γ równe 0,33 oraz prawdopodobieństwo równe 0,33. Z uwagi na silny wpływ prawdopodobieństwa na wynik, zostały przeprowadzone dodatkowe eksperymenty dla prawdopodobieństw 0, 0,1 oraz 0,2, aby dodatkowo sprawdzić, czy reguła „im mniejsze prawdopodobieństwo, tym lepiej" jest uniwersalna. Prawdopodobieństwo równe 0 spowodowoało znaczny skok w wartości wyników, co widać na wykresie 2.2, co sygnalizuje, że w tym problemie algorytm zachłanny spisuje się bardzo dobrze.



Rys. 2.1. Wykres przedstawiający uśrednione wyniki dla wszystkich kombinacji parametrówi



Rys. 2.2. Wykres przedstawiający uśrednione wyniki dla wybranej kombinacji parametrów ze zmiennym prawdopodobieństwem wyboru losowej polityki

2.5. Wnioski

Algorytm, jeśli zostanie nauczony wystarczająco dużą liczbą iteracji, radzi sobie dobrze z postawionym zadaniem, aczkolwiek zdarzają mu się „głupie" błędy, tzn. wjeżdżanie w ścianę, wysadzanie pasażera poza wyznaczonymi strefami, czy wracanie po własnych śladach. Ten problem zostaje zlikwidowany przy zastosowaniiu algorytmu zachłannego.