Sprawozdanie - SIECI SAMOUCZĄCE SIĘ

W ćwiczeniu sailor, wykorzystałem metodę Montecarlo i uczenia z krytykiem do rozwiązania zadania. Na samym początku tworzona jest losowa strategia –losowo wybierane są akcje spośród 4 akcji (przód, tył, lewo, prawo).  
Parametr **epsilon** odpowiada za wybór akcji. Określa prawdopodobieństwo z jakim akcja zostanie wybrana losowo zamiast według planowanej strategii. Parametr **alfa** jest współczynnikiem nagrody. Parametr **gamma** jest współczynnikiem dyskontowania.Następnie wykonywana jest funkcja environment zwracająca nagrodę i aktualny stan. Następnie stan jest ewaluowany.

Optymalizacja parametrów została zaimplementowana za pomocą 3 pętl (gamma alfa, epsilon). Poszukiwane parametry były w zakresie od 0 do 1 co 0.1. Każdy parametr był testowany. Wynikiem jest zestaw parametrów z najwyższym wynikiem

Osiągnięte zostały następujące wyniki:

|  |  |
| --- | --- |
| Nazwa pliku | [gamma, alfa, epsilon, wynik] |
| map\_small.txt | *[0.8, 0.4, 0.3, 2.329]* |
| map\_easy.txt | *[0.9, 0.2, 0.5, 6.692]* |
| map\_middle.txt | *[0.9, 0.2, 0.3, 6.9355]* |
| map\_big.txt | *[0.7, 0.5, 0.4, -1.3065]* |
| map\_spiral.txt | *[0.9, 0.2, 0.6, 178.7605]* |