

Plan wstępny

1:

Co to jest Reinforcement learning

Reinforcement learning czyli po polsku uczenie przez wzmocnianie

jest to metoda treningu uczenia maszynowego przez nagradzanie zachowań chcianych i/lub karanie zachowań nieodpowiednich. Metoda ta pozwala wyznaczyć optymalną politykę sterowania przez agenta w nieznanym mu środowisku.

2:

Q learning classic

Q learning jest typem ogólnie pojętego reinforcement learningu omawianego już wcześniej. W Q learningu posługujemy się podstawowymi pojęciami jak stan, akcja, agent, nagroda oraz kara.

W omawianym przypadku agent reprezentuje robota mobilnego, stan będzie pozycją agenta w środowisku a akcją będzie przemieszczenie się z jednego stanu do drugiego. Nagrodą jest wartość pozytywna - czyli zwiększająca

wartość danej akcji w danym, konkretnym stanie, a karą jest wartość negatywna.

Doświadczenie zdobywa się przez eksplorację oraz eksploatację przez agenta.

Eksploracją jest próbowanie nowych akcji w danym stanie a eksploatacją jest powtarzanie akcji już znanych.

W Q learningu nie jest wymagana znajomość prawdopodobieństw przejścia wszystkich akcji oraz stanów.

Wartości macierzy Q są kolejno obliczane przez następujące wyrażenie.

:

Gdzie

alfa i gamma od 0 do 1

alfa - jak bardzo odpowiedź jest ważna

gamma - patrzenie w przyszłość

3:

Problemy z Q learningiem

Biorąc pod uwagę przeszukiwaną przestrzeń z m stanami oraz n akcjami tabela Q będzie miała rozmiar m na n .

Poruszając się z jednego stanu do następnego agent wybiera akcję z największą wartością Q wśród n możliwych akcji.

Sprawia to że potrzebne jest $n-1$ porównań. Więc by zakualizować macierz Q, za każdym razem jest potrzebne $m \cdot (n-1)$ porównań.

Oznacza to że wraz z rozmiarem oraz skomplikowaniem środowiska, szczególnie w rzeczywistych rozwiązaniach, czas na znalezienie ścieżki wzrasta eksponencjalnie.

Drugim problemem jest całkowita losowość poruszania się agenta

skutkująca marnowaniem zasobów, wolniejszym zbieganiem się do ścieżki optymalnej oraz zajmującą znaczną ilość czasu.

jak już wcześniej mówiliśmy, akcja dla następnego stanu będzie stwierdzona przez największą wartość z macierzy Q.

Przez inicjalizację wartości na 0, na początku agent nie ma wyboru, musi być losowy. Zastępstwem tego może być inicjalizacja, w tym wypadku przez użycie algorytmu FPA.

4:

Co to jest FPA?

FPA, czyli flower polination algorithm

Zapylanie polega na przenoszeniu pyłków z jednego kwiatu na drugi kwiat,

Które można podzielić na z udziałem zwierząt przenoszących pyłki i bez.

90% korzysta ze zwierząt, zwane również zapylaniem krzyżowym, 10% samo-zapylanie.

Inspirowany zapylaniem naturalnym FPA Zaproponowane przez Yanga w 2012 ma 4 uproszczone zasady.

Dystrybucja Levy-iego $L \sim (\lambda * \gamma\{\lambda\}) * \sin(\pi * \lambda/2) / \pi * 1/s^{(1+\lambda)}$

Zapylanie globalne

Zapylanie Lokalne

5:

FPA jako lepsze zainicjowanie macierzy Q

Improved Decentralized Q-learning algorithm (IDQ) - został zmodyfikowany by oceniać różnicę między obecnym stanem a stanem końcowym

A - Q learning

B - IQ-FPA

In this regard, Arin and Rabadi integrated metaheuristic algorithm for randomized priority search with Q-learning, in order to minimize the search time for the optimal solution

Wang et al. proposed a backward Q-learning, combining the Sarsa algorithm and Q-learning [19].

The advantage of using Sarsa algorithm is that it takes a shorter time in convergence, as compared to Q learning. However,

Sarsa algorithm tends to be trapped in the local minimum, and this can be compensated by the hybridization with the Q-learning.

The backward Q-learning was applied in cliff walk, mountain car and cart-pole balancing control system,

where improvement in both learning time and finishing performance were observed.

Rakshit et al. combined the Q-learning as part of local refinement and differential evolution for global search in adaptive memetic algorithm