TP3 - Reinforcement Learning

Briac Six - Baptiste Bellamy

October 16, 2024

Abstract

Dans ce projet, nous avons implémenté et comparé deux algorithmes d' apprentissage par renforcement, Q-Learning et SARSA, pour résoudre le problème du jeu Taxi-v3 d'OpenAI Gym. Le but du jeu consiste à déplacer un taxi sur une grille 5x5, ramasser un passager à un emplacement donné, et le déposer à une destination spécifique tout en optimisant le nombre de mouvements. Le problème est un environnement de contrôle discret, idéal pour explorer des méthodes d'apprentissage tabulaire.

Notre objectif est d'analyser et de comparer les performances de ces deux algorithmes, à la fois en termes de vitesse d'apprentissage et de qualité des solutions obtenues. Nous avons mis en œuvre les deux algorithmes dans cet environnement, et nous présentons une analyse des résultats, ainsi que des vidéos illustrant le comportement de l'agent après entraînement.

Contents

1 Choix d'implémentation

Pour implémenter ces classes, nous avons suivi les indications vues en cours ainsi que la description des fonctions du TP. Nous avions cependant à déterminer les paramètres de chaque algorithme, soit :

- le taux d'apprentissage (α),
- γ (le facteur d'actualisation),
- ϵ (si besoin pour la politique ϵ-greedy).

Nous avons pour cela implémenté une fonction `search_best_parameters`, appelée sur chaque algorithme afin de trouver la combinaison des meilleurs paramètres. Nous avons alors trouvé que :

- Taux d'apprentissage (*learning rate*) : 0.75 (Q-Learning) et 0.8 (SARSA).
- Facteur de discount (γ) : 0.99, pour bien valoriser les récompenses futures.
- Epsilon (ϵ) : 0.1 pour Q-Learning avec epsilon fixe, et une valeur initiale de 1 pour epsilon scheduling, qui décroît progressivement.

Nous avons ensuite entraîné chaque agent pendant 550 epochs pour obtenir suffisamment de données.

2 Résultats et comparaison

Pour comparer les algorithmes, nous nous sommes basés sur plusieurs critères, dont le tableau résultant est présenté ci-dessous :

Critère	**QLearning**	**QLearningEpsilon**	**Sarsa**
Max Value	14.0	15.0	15.0
MaxValue Epoch	244	314	42
Mean Last 100	1.35	4.40	6.79
First positive rewards (epoch)	13	40	23

Table 1: Comparaison des performances des algorithmes

SARSA se démarque par sa rapidité à converger et par une performance stable sur les 100 derniers épisodes, mais son exploration peut être plus limitée. QLearningEpsilon offre un bon compromis entre exploration et exploitation, atteignant une bonne performance à long terme, mais nécessitant plus de temps pour converger. QLearning classique, avec un epsilon fixe, est plus rapide à obtenir une première récompense positive mais offre des performances globales inférieures à celles des deux autres algorithmes. Dans l'ensemble, SARSA semble le plus performant en termes de stabilité et de rapidité, tandis que QLearningEpsilon est plus performant que QLearning classique, tout en prenant plus de temps à converger. Pour appuyer nos résultats, nous avons réalisé des vidéos de nos algorithmes en action en utilisant la bibliothèque Gymnasium, afin de visualiser et illustrer leur performance dans l'environnement Taxi-v3.