

Rapport TP 3 Reinforcement Learning

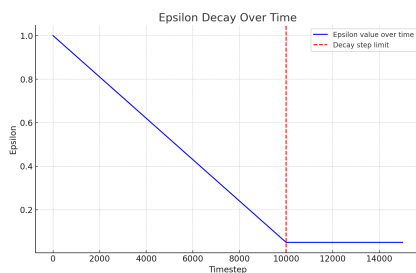
Dans ce projet, nous avons exploré plusieurs algorithmes de Reinforcement Learning, notamment le Q-Learning et Sarsa pour entraîner des agents à jouer au jeu Taxi d'openAI Gym.

Q Learning Agent

Nous avons implémenté le Q Learning Agent dans un premier temps avec les hyper paramètres initiaux ($\epsilon = 0.25$, $\text{learning_rate} = 0.5$, $\gamma = 0.99$). Ce premier test n'a pas donné de bon résultats, avec une récompense toujours négative après 1000 épisodes. Après avoir étudié nos résultats, nous avons remarqué que notre agent était trop souvent en phase d'exploration, donc nous avons décidé de réduire la phase d'exploration en diminuant epsilon ($\epsilon = 0.1$) et nous avons obtenu de bien meilleurs résultats avec une récompense positive des 500 épisodes.

Q Learning Agent Epsilon Scheduling

Pour implémenter le Q Learning Agent Epsilon Scheduling, nous avons utilisé le calcul suivant:
$$\epsilon = \epsilon_{\text{start}} - (\epsilon_{\text{start}} - \epsilon_{\text{end}}) * (\text{timestep} / \epsilon_{\text{decay_steps}})$$



Nous avons obtenu de meilleurs résultats avec cet algorithme car il permet de mieux gérer la balance entre exploration et exploitation.

Sarsa

Pour implémenter l'algorithme Sarsa nous avons simplement suivi les calculs et avons obtenu des résultats satisfaisants.

Comparaison des différents algorithmes

	Reward (ep = 500)	Reward (ep = 1000)
Q Learning Agent	1.27	2.2
Q Learning Agent Eps Scheduling	3.4	4.97
Sarsa	6.35	6.73

En conclusion, notre étude a montré que l'algorithme Sarsa offre une meilleure performance dans notre environnement, en raison de sa capacité à mieux ajuster la politique au fur et à mesure de l'apprentissage.

Toutes les vidéos des agents peuvent être trouvées dans le dossier "videos/".