

Elias Tebbani

Anthony Belval

SCIA 2025

Rapport RL TP3

Qlearning:

Dans ce fichier, nous avons implémenté le code en suivant les consignes en commentaire il n'y pas de choix particulier d'implémentation.

Qlearning_eps_scheduling :

Nous avons ici aussi suivit les consignes d'implémentation notre choix pour la formule utilisée pour ajuster l'epsilon est la suivante :

epsilon =

$\max(\text{epsilon_end}, \text{epsilon_start} - (\text{epsilon_start} - \text{epsilon_end}) * \text{timestep} / \text{epsilon_decay_steps})$

À chaque time step, la valeur d'epsilon diminue progressivement jusqu'à atteindre un plancher fixé par epsilon_end. Cela garantit que l'agent explore beaucoup au début et exploite davantage ses connaissances plus tard dans l'entraînement, tout en maintenant une petite part d'exploration pour éviter de rester coincé dans des solutions sous-optimales.

Sarsa :

Ici, nous avons implémenté le code en suivant les consignes en commentaire il n'y pas de choix particulier d'implémentation.

Taxi:

Nous avons ajusté certains paramètres de l'agent Q-Learning. Le taux d'apprentissage a été fixé à 0,5 pour favoriser un apprentissage plus rapide, car un taux trop bas freinait la convergence. L'epsilon a été réduit à 0,05 afin de diminuer la probabilité d'actions aléatoires, qui était de 25 % auparavant, ce qui compliquait l'apprentissage. Le gamma est resté à 0,99 pour privilégier les récompenses à long terme. Enfin, le nombre maximum d'actions (t_max) a été diminué à 200 pour permettre à l'agent de compléter ses épisodes plus rapidement. Pour les autres modèles nous n'avons pas changé les paramètres. Pour enregistrer les vidéos des agents en action, nous avons simplement utilisé l'outil intégré de la Library gym.