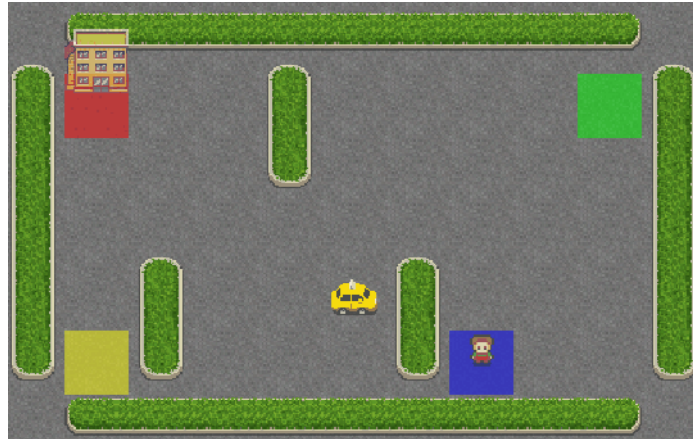


# Rapport TP3

## Reinforcement Learning



EPITA - [SCIA] - 2025

*Octobre 2024*

Sami CARRET ( *sami.carret@epita.fr* )  
Florine KIERAGA ( *florine.kieraga@epita.fr* )

# 1 Rapport

## 1.1 Contexte

Pour ce TP, nous avons testé 3 méthodes de Reinforcement Learning pour le modèle Taxi-V3 de OpenAI. Ces trois méthodes d'entraînement sont :

- Q-Learning
- Q-Learning avec *epsilon scheduling*
- SARSA

Le modèle Taxi-V3 consiste en une carte sur laquelle se trouve un taxi. La carte contient quatre carrés colorés avec une personne sur l'un d'eux et un hôtel sur un autre. L'objectif du TP est d'entraîner le taxi à récupérer la personne et à la ramener à l'hôtel de la manière la plus efficace possible.

## 1.2 Implementation des algorithmes

Pour l'implémentation des algorithmes, nous avons simplement appliqué les formules données lors du cours et dans les commentaires du code. Chaque entraînement a été effectué sur 1000 épisodes.

Nous avons ensuite cherché à trouver les meilleurs hyperparamètres avec un GridSearch pour l'algorithme de Q-Learning ( les paramètres : *Learning Rate*, *Epsilon*, *Gamma*) [Annexe A.1.1]. Cette recherche a donné les résultats suivants : **Learning Rate** : 0.9, **Epsilon** : 0.10 et **Gamma** : 0.81.[Annexe A.1.2] Le GridSearch a été effectué sur 250 épisodes pour chaque combinaison d'hyperparamètres pour avoir un compromis entre précision et vitesse.

Nous avons réutilisé les valeurs trouvées par le GridSearch dans tous les algorithmes.

Pour chaque algorithme, nous avons généré des vidéos de l'entraînement à l'épisode 0, 250, 500, 750 et 1000 (l'épisode finale, le résultat). Ces vidéos sont présentes sur le repo du tp dans le dossier *videos*.

## 1.3 Résultats

Après avoir exécuté le code, on peut voir sur le graphique des learning curves [Annexe A.2] que le Q-Learning avec *epsilon scheduling* atteint le plus rapidement la maximisation de sa récompense. Ensuite, c'est l'algorithme Q-Learning normal et puis enfin l'algorithme SARSA.

En conclusion, la méthode Q-Learning est plus efficace que SARSA pour le problème Taxi-V3 de OpenAI, et elle devient nettement plus performante avec l'ajout du *epsilon scheduling*.

## A Annexes

### A.1 Grid Search

#### A.1.1 Résultats du Grid Search

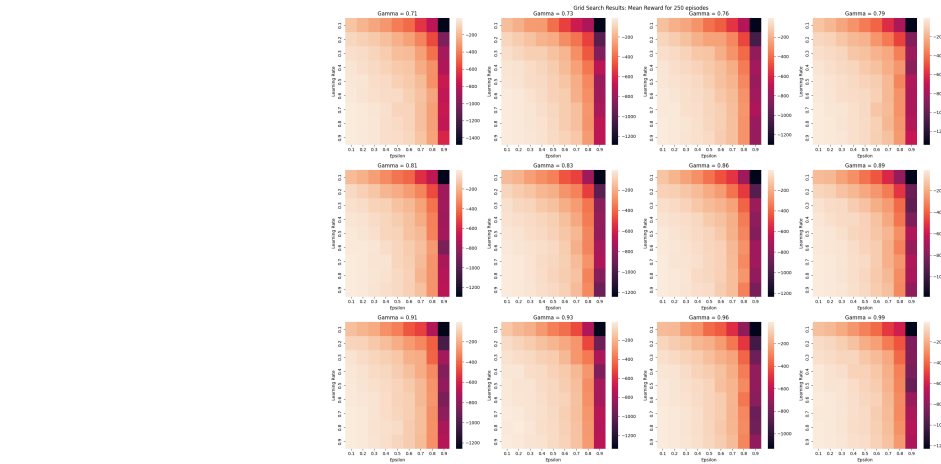


FIGURE 1 – Résultats du Grid Search pour l'agent Q-Learning

#### A.1.2 Meilleurs valeurs du Grid Search

```
Best hyperparameters: Learning Rate = 0.90, Epsilon = 0.10, Gamma = 0.81
```

FIGURE 2 – Meilleurs valeurs du Grid Search

## A.2 Learning Curves

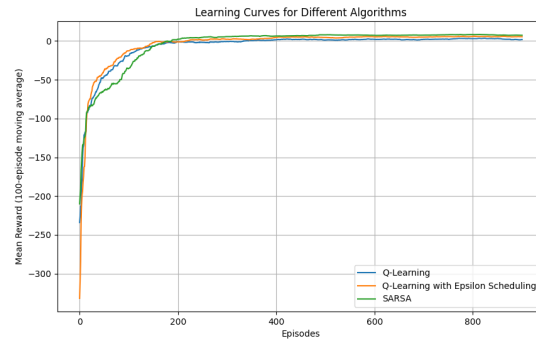


FIGURE 3 – Learning Curves pour les différents algorithmes