Virgile Hermant
Lylian Cazale

TP3 RL :
SARSA & Q-Learning pour Taxi-v3

I. Implémentation

A. Q-Learning

Pour l'implémentation de Q-learning, nous avons implémenté le fait de trouver la meilleure action, puis de l'utiliser avec une probabilité de 1 - epsilon. Sinon, nous choisissons une action au hasard. Le système d'update, qui permet à notre modèle d'apprendre, va mettre à jour les meilleures actions en fonction des rewards que nous avons obtenus.

B. Q-Learning scheduling

L'implémentation de la version scheduling de Q-Learning change uniquement sur les actions qui vont être faites lors de la get_value. Au fur et à mesure des utilisations, nous allons réduire epsilon. Ce qui aura pour effet de réduire les chances de choisir une action au hasard. Cela permettra, à terme, de converger vers une situation.

C. SARSA

La principale différence entre SARSA et Q-Learning réside dans la manière dont les Q-values sont mises à jour. Q-Learning met à jour les Q-values indépendamment de l'action réellement choisie par l'agent, tandis que SARSA les ajuste en tenant compte de l'action effectivement sélectionnée. Q-Learning tend à converger plus rapidement vers une solution, mais peut être instable dans certaines situations. En revanche, SARSA converge plus lentement, mais offre une plus grande stabilité à travers divers scénarios.

II. Résultats

Pour ce qui est des résultats, nous avons observé 3 modèles qui sont assez similaires, bien que les Q-Learning aient convergé légèrement plus rapidement, mais cela est rendu invisible au bout de quelques étapes supplémentaires. Également, les Q-learnings ont été instable au début des simulations lors de tentatives risquées qui leur a fait avoir de très mauvais résultats sur certains épisodes. Sarsa, quant à lui, à converger moins rapidement, mais sans grands écarts au fur et à mesure des épisodes.

Comme prévu, sarsa a été plus lent mais plus stage que Q-Learning. Les chemins d'apprentissage des algorithmes étaient différents mais ils sont tous arrivés au même point. L'utilisation des algorithmes dépend de l'utilisation que l'on veut en faire. Si l'on veut quelque chose de stage et de sur pendant l'entraînement, SARSA sera le meilleur choix. D'un autre côté, si la stabilité n'est pas recherchée mais justement plus l'exploration et/ou la convergence rapide, Q-Learning peut avoir ses chances.

Virgile Hermant
Lylian Cazale

Paramètres :

Nombre d'epoch : 1000

- **Q-Learning**
  - Learning Rate = 0.5
  - Epsilon = 0.1
  - Gamma = 0.99
- **Q-LearningScheduling**
  - Learning Rate = 0.5
  - Epsilon = 0.25
  - Gamma = 0.99
- **SARSA**
  - Learning Rate = 0.5
  - Gamma = 0.99

Tous les GIF des vidéos sont disponible dans le dossier videos/