

Analyse Q-Learning vs SARSA TP-3 RL

Abel ANDRY, Emile MERLE

20 octobre 2024

1 Analyse Q-LEARNING

Choix d'implémentation : Pour Q-Learning, nous avons décidé de faire les choses suivantes :

- Nous utilisons Q-Learning avec Epsilon Scheduling pour les observations suivantes.
- Nous implémentons du Epsilon-decay en plus de l'Epsilon-greedy pour réduire les chances d'un choix aléatoire au fur et à mesure que les actions sont choisies, sans pour autant les enlever complètement.

Performances :

- Temps de convergence : 2.2 secondes
- Récompense finale : + 4.92
- Progression des récompenses : -1204.0 → -403.08 → -26.87 → -0.36 → 1.83 → 3.0 → 4.79 → 4.48 → 5.3 → 4.92

Caractéristiques principales :

- Courbe d'apprentissage moins régulière que SARSA
- Comportement plus libre (et donc propice à l'erreur) dû à l'approche "off-policy"
- Stabilisation progressive après 400 épisodes
- Convergence plus rapide mais moins stable que SARSA

La méthode du Q-Learning propose une approche moins fiable mais plus rapide à converger, ce qui peut donner des résultats variables. Cependant, ce genre d'approche peut être une bonne idée lorsque l'on recherche à avoir des résultats rapidement, quitte à ne pas avoir de stratégie mise en place.

2 Analyse SARSA

Performances :

- Temps de convergence : 2.8 secondes
- Récompense finale : +5.92
- Progression des récompenses : -1294.0 → -389.08 → -105.9 → -39.8 → -14.25 → -4.86 → 0.69 → 6.1 → 4.79 → 5.92

Caractéristiques principales :

- Courbe d'apprentissage plus régulière que Q-Learning
- Comportement conservateur dû à l'approche "on-policy"
- Stabilisation progressive après 400 épisodes
- Convergence plus lente mais plus stable que Q-Learning

En conclusion, SARSA présente une approche plus conservative mais plus fiable, privilégiant la stabilité à la vitesse de convergence. Cette caractéristique en fait une option pertinente pour les applications nécessitant une progression d'apprentissage régulière.

3 Conclusion

En somme, nous pouvons conclure que les deux approches offrent des opportunités différentes qui peuvent s'appliquer à différents cas de figure ou de nécessité d'applications. Dans le cas où l'on cherche à avoir des résultats fiables, bien que plus lent, SARSA est sans doute la meilleure approche. Si l'on cherche à avoir une

vitesse de convergence rapide avec comme contre-partie une approche plus aléatoire, le Q-Learning est alors plus adapté.

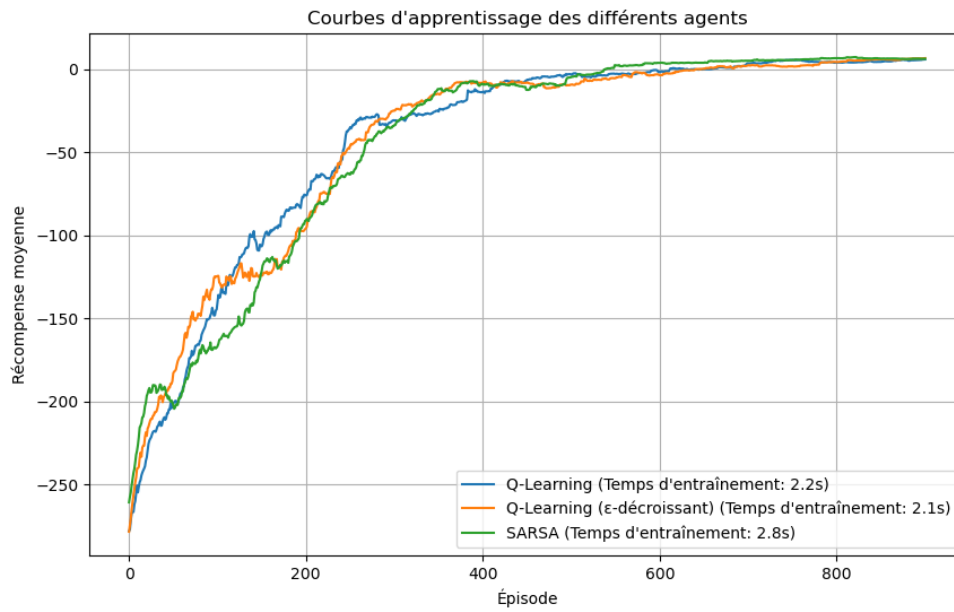


FIGURE 1 – Courbes d'apprentissage des différents agents