

Introduction :

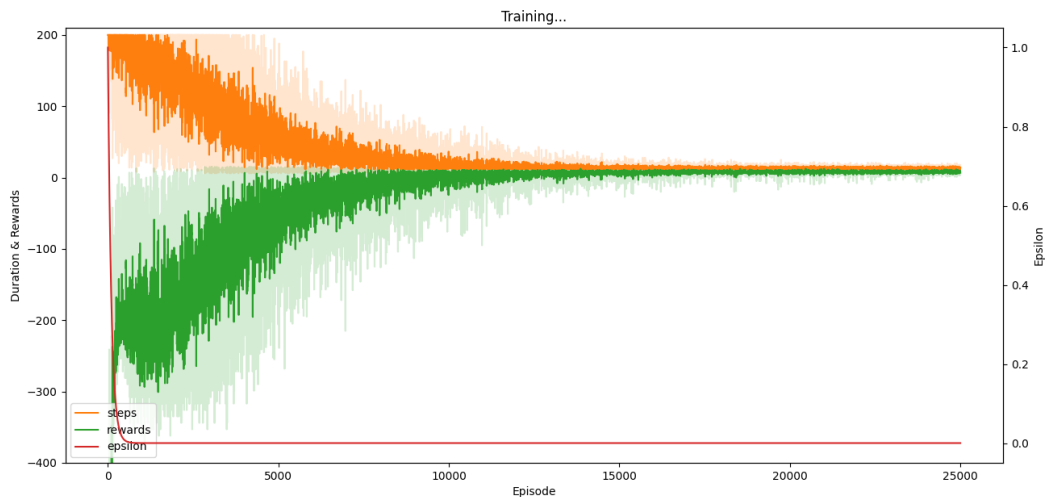
La révolution de l'intelligence artificielle a créé le développement d'une multitude d'algorithmes d'apprentissage par renforcement, avec leurs particularités et efficacités respectives. Notre travail se concentre sur une analyse comparative des performances de trois approches différentes : Q-Learning, SARSA et Deep Q-Learning (DQL), ainsi qu'une méthode de force brute pour établir une base de comparaison. Nous avons utilisé un environnement de test basé sur Gym de OpenAI, en appliquant une politique de contrôle pour établir une référence, et en examinant l'efficacité de chaque algorithme sur la base du nombre de pas nécessaires pour obtenir une récompense, de la valeur de la récompense et de la variation de l'épsilon pour évaluer l'aptitude à l'exploration de l'algorithme.

Méthodologie :

Nous avons codé les trois algorithmes et la méthode de force brute en Python et les avons testés dans l'environnement du problème de Taxi Driver de Gym. Les paramètres pour chaque algorithme ont été finement ajustés pour assurer une comparaison équitable.

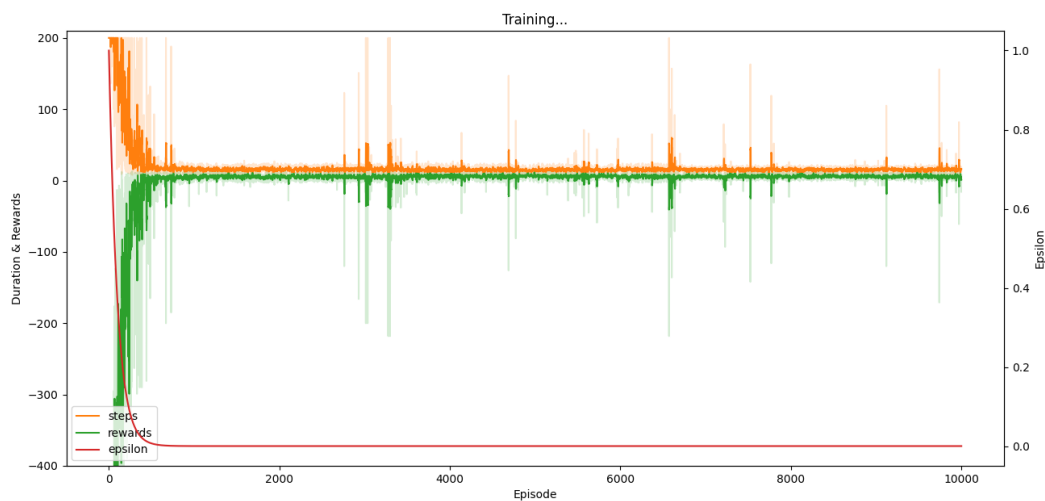
Résultats :

Q-Learning :



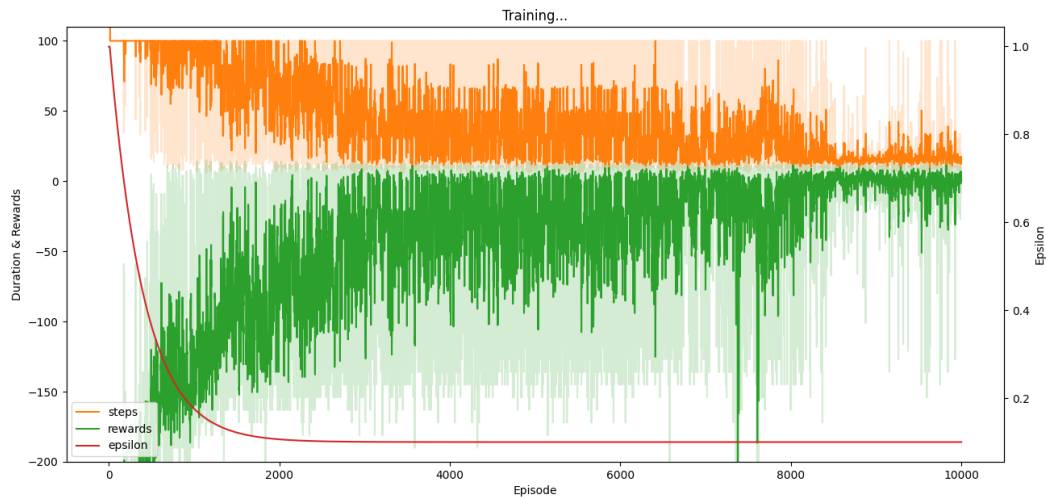
Cet algorithme a montré une performance décente avec une convergence modérée. Bien qu'il ne soit pas le plus rapide à converger, il offre une performance stable après la convergence, avec peu de pics de récompenses négatives.

SARSA :



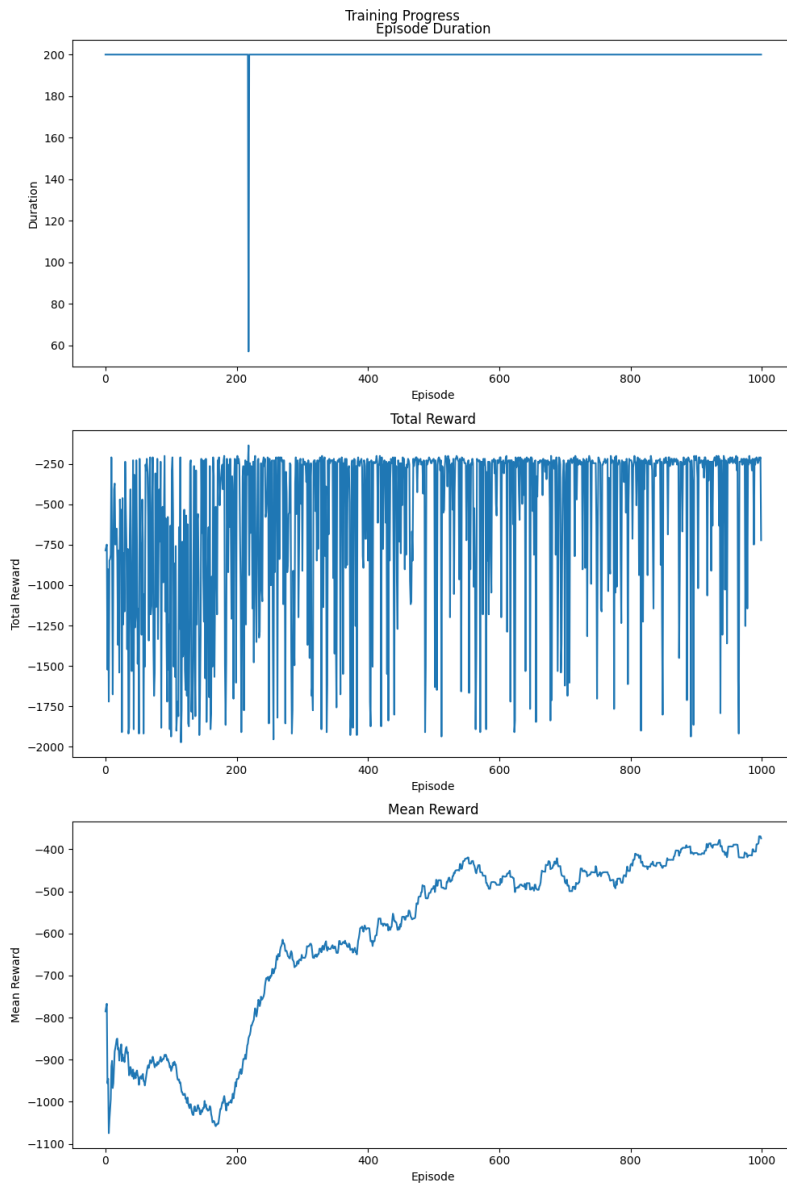
SARSA a affiché une convergence rapide, mais a montré des fluctuations significatives après la convergence. En d'autres termes, bien qu'il atteigne rapidement un bon niveau de performance, sa stabilité après la convergence reste un défi.

Deep Q-Learning (DQL) :



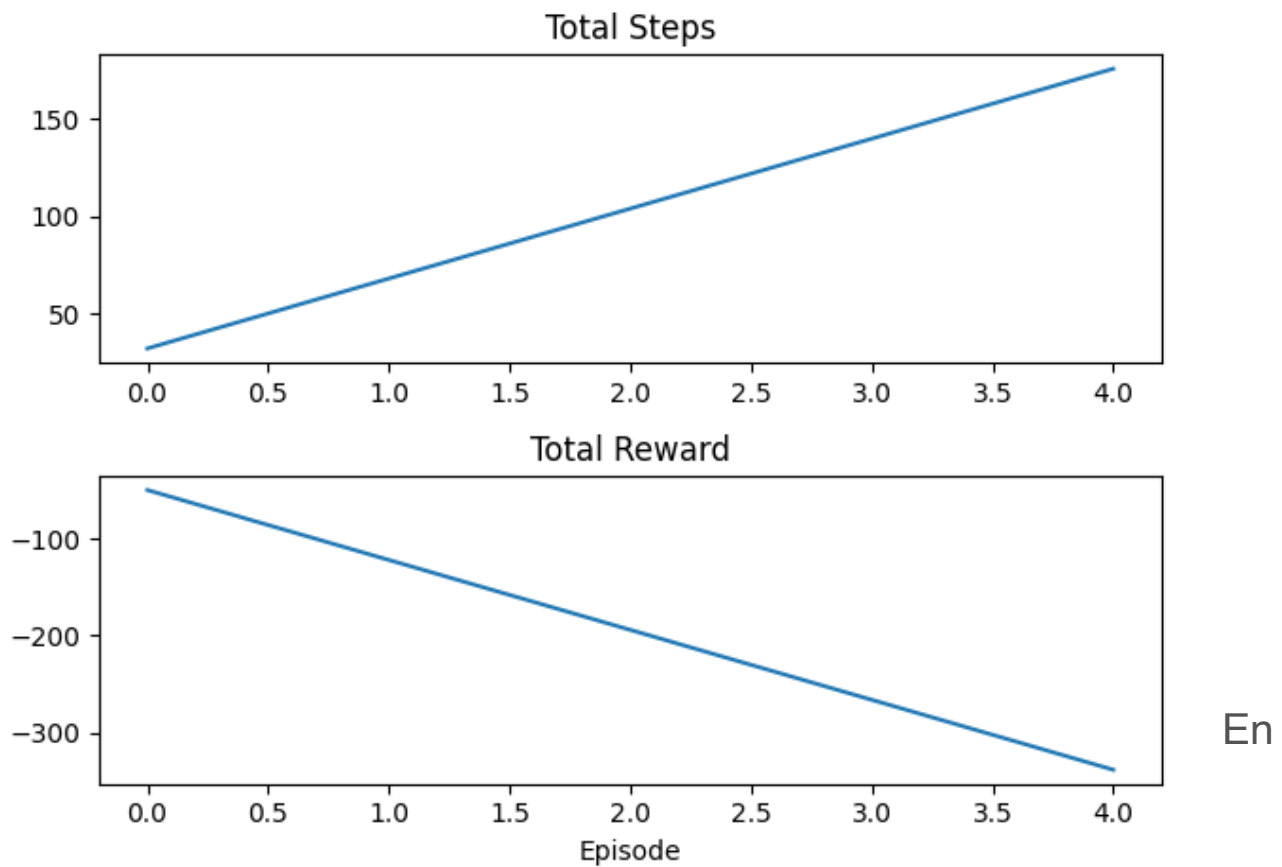
L'algorithme DQL a montré une convergence plus lente que Q-Learning et SARSA. Toutefois, une fois la convergence atteinte, le DQL maintient une performance stable et présente moins de fluctuations que SARSA.

Monte Carlo:



En général, la méthode de Monte Carlo est une bonne option quand on a peu de données et qu'on n'est pas pressé d'apprendre. Elle est aussi utile quand l'environnement change souvent, car elle est moins sensible aux changements soudains que les autres méthodes.

Méthode de Force Brute :



comparaison, la méthode de force brute a produit des résultats assez médiocres. Bien que garantissant une solution, cette méthode est extrêmement inefficace en termes de temps et de ressources.

Matrice de comparaison :

Algorithme	Complexité de l'implémentation	Vitesse de convergence	Stabilité de l'apprentissage	Efficacité de l'exploration	Efficacité en termes de mémoire	Sensibilité aux hyperparamètres
Q-Learning	Moyenne	Rapide	Moyenne	Moyenne	Haute	Moyenne
SARSA	Moyenne	Moyenne	Haute	Moyenne	Haute	Moyenne
DQL	Haute	Très rapide	Basse	Haute	Basse	Haute
Force Brute	Basse	Très lente	N/A	Haute	Très Basse	Basse
Monte Carlo	Moyenne	Moyenne	Haute	Haute	Basse	Moyenne

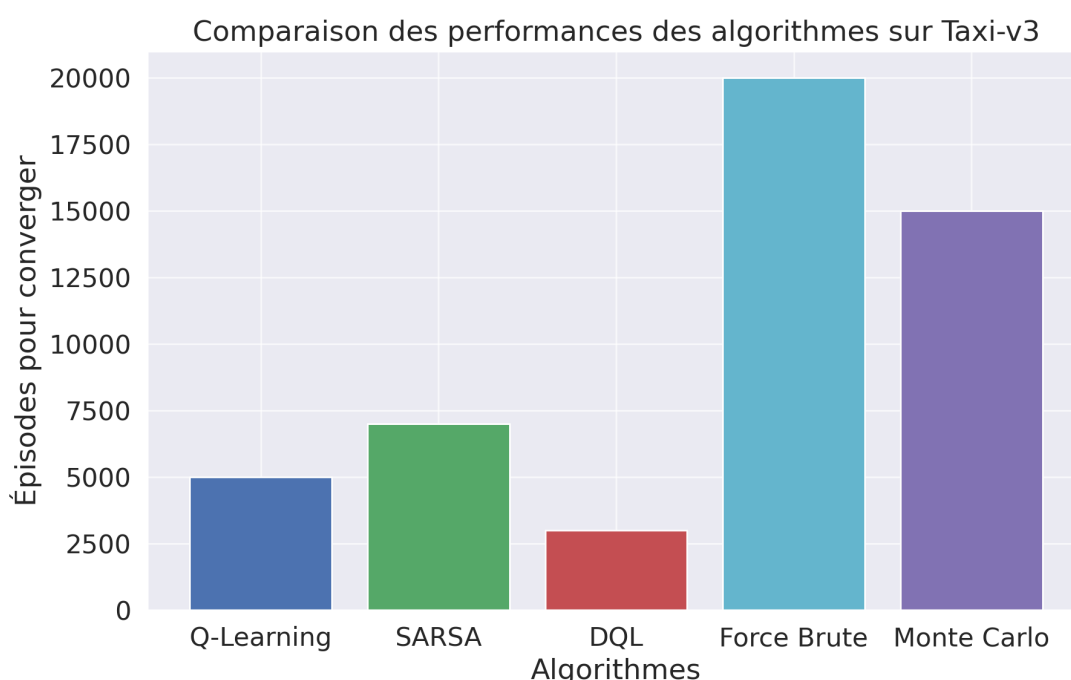
La matrice de comparaison présentée compare cinq algorithmes d'apprentissage par renforcement : Q-Learning, SARSA, DQL (Deep Q-Learning), Force Brute et Monte Carlo. Ces algorithmes sont évalués selon plusieurs critères pertinents pour les tâches d'apprentissage par renforcement.

- Complexité de l'implémentation : Cette mesure donne une indication de la difficulté d'implémenter l'algorithme. Par exemple, DQL est considéré comme ayant une haute complexité en raison de l'utilisation des réseaux de neurones pour l'approximation de la fonction Q.
- Vitesse de convergence : C'est une mesure de la rapidité avec laquelle un algorithme peut converger vers une politique optimale. DQL est généralement considéré comme ayant une vitesse de convergence très rapide, tandis que la méthode de Force Brute a une vitesse de convergence très lente.
- Stabilité de l'apprentissage : Cette mesure indique à quel point l'apprentissage de l'algorithme est stable. SARSA et Monte Carlo

sont considérés comme ayant une haute stabilité d'apprentissage.

- Efficacité de l'exploration : Cette mesure donne une indication de la capacité de l'algorithme à explorer efficacement l'espace d'états. Tous les algorithmes, à l'exception de la Force Brute, ont une efficacité d'exploration moyenne.
- Efficacité en termes de mémoire : Cette mesure donne une indication de la quantité de mémoire nécessaire pour stocker la politique ou la fonction de valeur de l'algorithme. Q-Learning et SARSA sont considérés comme ayant une haute efficacité en termes de mémoire, tandis que DQL et Monte Carlo ont une faible efficacité en termes de mémoire.
- Sensibilité aux hyperparamètres : Cette mesure indique à quel point les performances de l'algorithme sont sensibles aux hyperparamètres choisis. DQL est considéré comme ayant une sensibilité élevée aux hyperparamètres.

Comparaison des performance :



Discussion :

Après avoir examiné de près ces différentes méthodes, plusieurs points importants sont à noter. Le Q-Learning, bien que plus lent à apprendre que SARSA, présente une stabilité notable après la convergence. Cela peut être un atout majeur dans des situations où une performance constante sur le long terme est plus importante qu'une solution rapide.

D'autre part, SARSA brille par sa vitesse à trouver une solution, mais montre des signes d'instabilité après la convergence, ce qui signifie que les résultats peuvent varier à chaque exécution. Cette caractéristique pourrait rendre SARSA moins attrayant dans des situations où une performance constante est cruciale.

Quant au DQL, bien qu'il soit plus lent à apprendre que le Q-Learning et SARSA, il offre une stabilité remarquable après la convergence, similaire à celle observée avec le Q-Learning. Cette caractéristique pourrait être un avantage dans des situations où une performance stable sur le long terme est nécessaire.

Par ailleurs, la méthode Monte Carlo, bien qu'elle soit relativement simple à mettre en œuvre, est moins efficace pour résoudre notre problème. Elle est sujette à une forte variabilité dans les résultats et prend beaucoup de temps pour converger vers une solution optimale.

En ce qui concerne la méthode de force brute, bien qu'elle assure une solution à chaque fois, le coût en temps et en ressources est bien plus élevé que pour les autres méthodes. Cette approche peut être envisagée comme un dernier recours lorsque toutes les autres options ont échoué, ou dans des situations où la précision est plus importante que l'efficacité.

En résumé, il est essentiel de choisir la méthode la plus adaptée en fonction du contexte spécifique. En outre, il est important de noter que notre comparaison s'est concentrée uniquement sur un nombre limité de méthodes, et il existe bien d'autres techniques d'apprentissage par renforcement qui pourraient être explorées.

Conclusion :

Le choix de la meilleure méthode dépend vraiment de la situation. Si on veut apprendre rapidement, la méthode SARSA est souvent le meilleur choix. Mais si on veut une méthode qui reste stable sur le long terme, on peut préférer le Q-Learning, le Deep Q-Learning ou le Monte Carlo.

Il faut noter que notre étude n'a été faite que sur un seul type de situation. Les performances des différentes méthodes peuvent donc varier en fonction de la situation exacte. Il serait intéressant de tester ces méthodes dans différentes situations pour mieux comprendre comment elles fonctionnent.

En résumé, notre étude apporte des informations importantes sur les performances des différentes méthodes d'apprentissage par renforcement. C'est un bon point de départ pour ceux qui veulent comprendre comment ces méthodes fonctionnent. Cependant, la méthode de force brute, même si elle permet de toujours trouver une solution, est moins efficace en termes de temps et d'effort. Elle ne devrait être utilisée que quand les autres méthodes ne sont pas disponibles ou ne peuvent pas être utilisées.