



Q-LEARNING

Le machine learning avec apprentissage par renforcement



Présenté par:

- HABIBI Mohamed

Encadré par: Mr A.ghazdali



SOMMAIRE

01

DEFINITIONS

1. Reinforcement learning
2. Q-learning

02

TERMES IMPORTANTS

1. Etat
2. Action
3. Récompense
4. épisode
5. Q-values
6. diff temporelle

03

AGIR

1. exploitation
2. exploration

SOMMAIRE

04

EQUATION
BELLMAN

05

ALGORITHME
Q-LEARNING

06

IMPLEMENTATION
SUR PYTHON

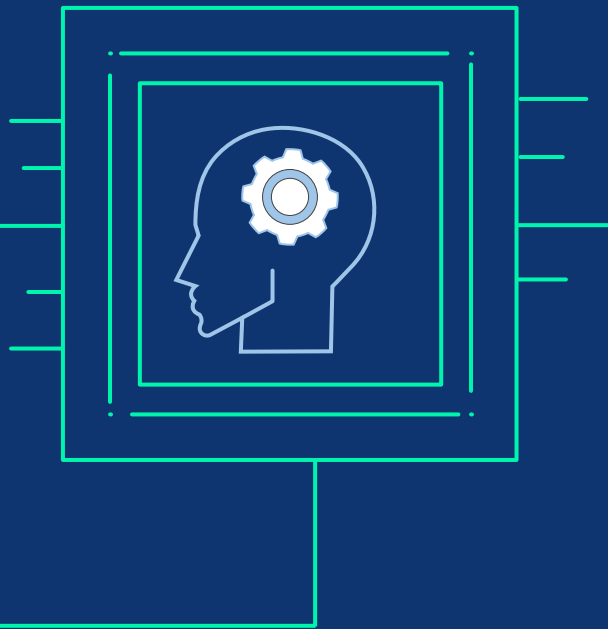
Definitions

Qu'est ce que le reinforcement learning ?

Le reinforcement learning est une méthode de machine learning dont l'objectif est de permettre à un agent, placé dans un environnement interactif, de choisir des actions maximisant des récompenses quantitatives. L'agent fait des essais et améliore sa stratégie d'action en fonction des récompenses fournies par l'environnement.



Definitions



Qu'est-ce que le Q-learning ?

Le Q-learning est un algorithme d'apprentissage par renforcement qui cherche à trouver la meilleure action à entreprendre compte tenu de l'état actuel.

Le "Q" de Q-learning est synonyme de qualité. La qualité représente l'utilité d'une action donnée pour obtenir une récompense future.



TERMES IMPORTANTS

ETAT

"S", représente la position actuelle d'un agent dans un environnement.

ACTION

l'action, "A", est le pas effectué par l'agent lorsqu'il se trouve dans un état particulier.

RÉCOMPENSE

Pour chaque action, l'agent recevra une récompense positive ou négative.

ÉPISODE


Lorsqu'un agent se retrouve dans un état de terminaison et ne peut pas effectuer de nouvelle action.

Q-VALUES

Utilisé pour déterminer la qualité d'une action, "A", prise à un état particulier, "S". $Q(A, S)$.

DIFFÉRENCE TEMPORELLE

Une formule utilisée pour trouver la valeur "Q" en utilisant la valeur de l'état et de l'action actuels et précédentes .



AGIR



EXPLOITATION

Cette manière consiste à utiliser le Q-table comme référence et à visualiser toutes les actions possibles pour un état donné. L'agent sélectionne alors l'action basée sur la valeur de récompense maximale de ces actions.



EXPLORATION

La deuxième manière consiste à agir de façon aléatoire. C'est ce qu'on appelle l'exploration. Au lieu de sélectionner des actions en fonction de la récompense future maximale, nous choisissons une action au hasard.

EQUATION DE BELLMAN

Q-Value actuel

récompense

maximum de la récompense
future

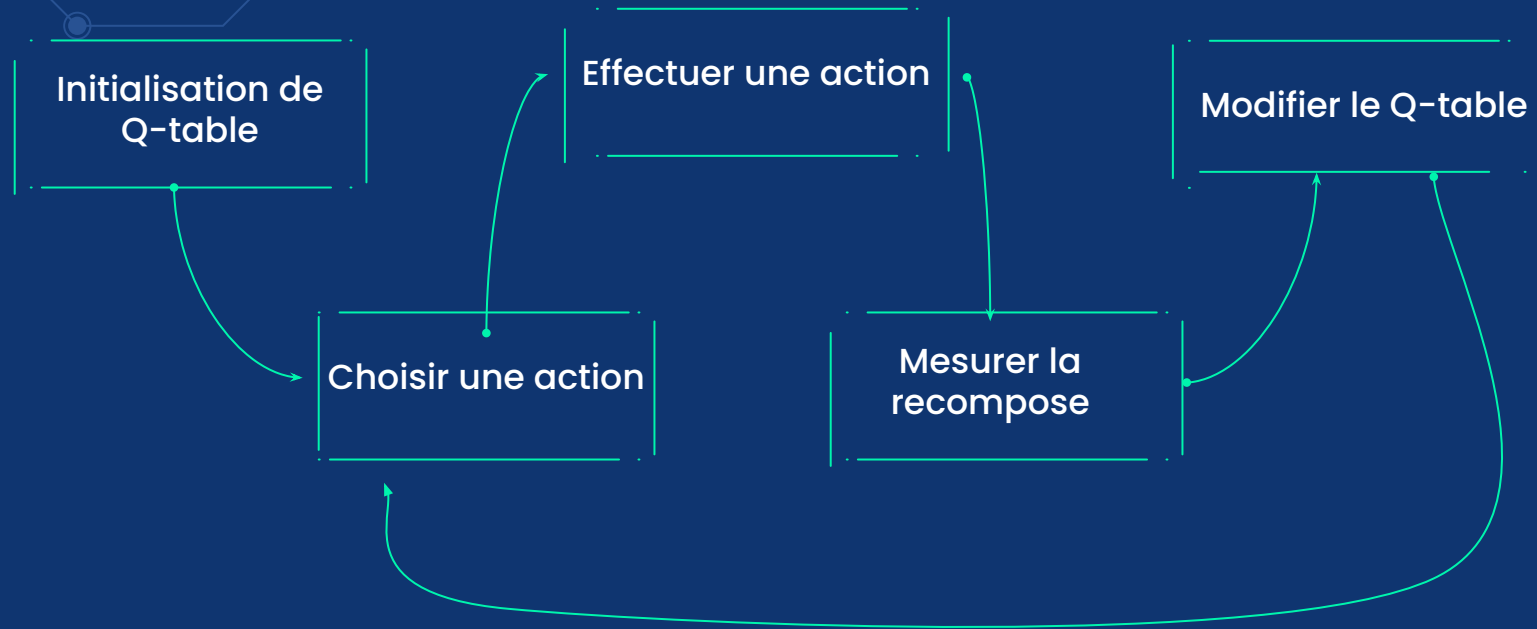
$$\text{New } Q(S,A) = Q(S,A) + \alpha \times [R(S,A) + \gamma \times \text{Max } Q(S',A) - Q(S,A)]$$

Taux d'apprentissage

Facteur d'actualisation



DIAGRAMME Q-LEARNING



IMPLEMENTATION PYTHON

