IA

hackA tech

Shake science. Shape innovation.

#decision

#apprentissage

#essaierreur

APPRENTISSAGE PAR RENFORCEMENT
Les bienfaits de l'expérience

Inria

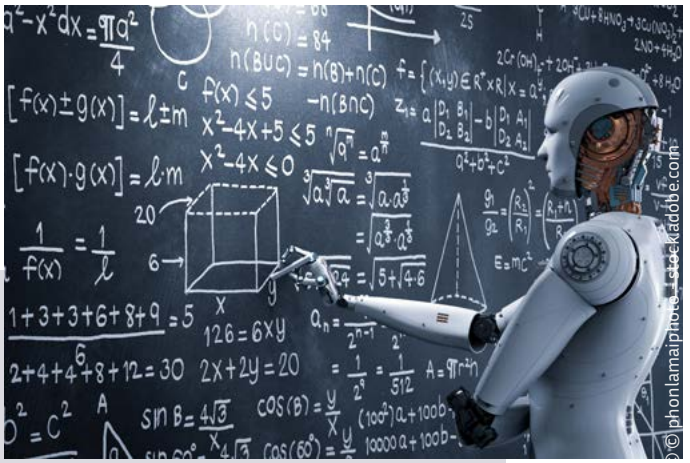
CARACTÉRISTIQUES

L'apprentissage par renforcement (*renforcement learning*) dit RL est une branche de l'Intelligence Artificielle qui permet à un programme d'apprendre par l'expérience.

Pour cela, un algorithme (*agent*) interagit avec un environnement incertain ou inconnu pour apprendre quel comportement avoir, afin de réaliser une tâche le mieux possible (*maximiser le niveau de réussite*). Une des particularités de cet algorithme est qu'il apprend tout seul. Lorsque l'algorithme est efficace, les résultats sont meilleurs que ceux obtenus par des humains sur les mêmes tâches. L'expert humain peut donc s'inspirer de ce qui a été appris.

Besoins nécessaires :

Pour fonctionner, il faut mettre en place un environnement (*réel, simulateur ou modèle*) pour extraire des cas d'apprentissage. Il est également nécessaire de pouvoir extraire un signal qui note le niveau de réussite de l'algorithme selon le résultat attendu.



USE CASES

Les domaines d'application sont nombreux :

- **Robotique :** déplacement autonome d'une machine,
- **Santé :** stratégie de soins
- **Gestion complexe :** pilotage des centrales électriques, gestion d'investissement en bourse,
- **Découverte de stratégie :** jeu de Go,
- **Logistique :** optimisation de trajet.

QUELS AVANTAGES ?

- Pas besoin d'expert
- Pas besoin de données
- Foctionne avec un environnement réel ou simulé

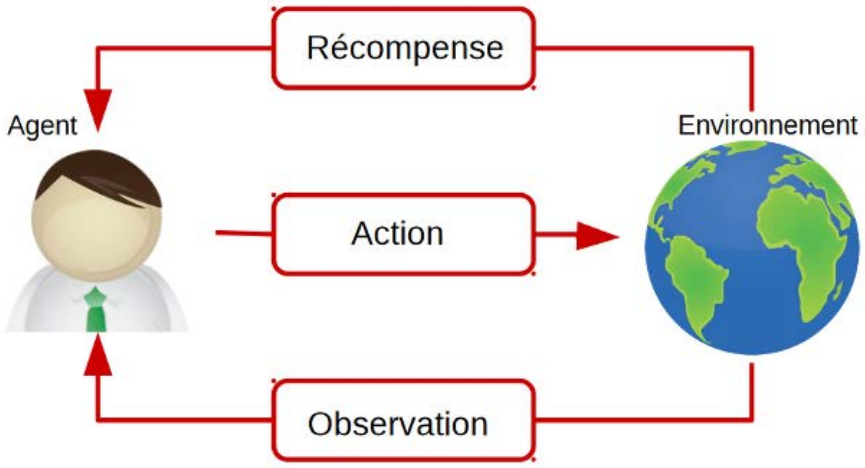
FICHE IDENTITÉ

- Langage de programmation: : Python (le plus utilisé), R, C, C++
- Outils : scikit-learn, tensorflow ou pytorch
- Caractéristiques : Modélisation de la tâche à optimiser par un problème de décision de Markov / Gestion du dilemme exploration / exploitation / Essai-erreur pour découvrir les meilleures solutions
- Équipe projet : LACODAM

FONCTIONNALITÉS GÉNÉRIQUES

- **Exploration :** l'agent essaye des actions variées pour augmenter sa connaissance de l'environnement afin de trouver la nouvelle meilleure action.
- **Exploitation :** l'agent choisit la meilleure action selon ses connaissances actuelles pour maximiser sa récompense. L'objectif est de trouver un compromis pour découvrir les meilleures actions, sans sacrifier trop de récompenses.

À partir d'un environnement, d'une liste d'actions possibles et d'un objectif à résultat quantifiable, le système apprend le comportement qui maximisera l'objectif. L'agent observe l'environnement pour savoir dans quel état il se trouve. En fonction de sa politique (expériences précédentes ou connaissances) et d'une possibilité d'exploration, une action est choisie. Suite à cette action, l'environnement se met à jour et envoie à l'agent la récompense (positive ou négative) correspondant à son action. L'agent peut alors observer l'environnement pour découvrir son nouvel état. Il utilise l'ancien état, l'action choisie, la récompense et le nouvel état pour mettre à jour sa politique. L'agent recommence une nouvelle action à partir du nouvel état et de ses connaissances.



READ ME

http://www.incompleteideas.net/book/RLbook2018.pdf