

#Décision

#Apprentissage

#AgentAutonome



Reinforcement learning

Les bienfaits de l'exploration

SOLUTION

L'apprentissage par renforcement est une branche de l'apprentissage automatique qui permet à un programme d'apprendre par l'expérience. Le principe est le suivant : un algorithme (agent) interagit avec un environnement incertain ou inconnu pour apprendre un comportement optimal, afin de réaliser une tâche le mieux possible.

Une des particularités de cette famille d'algorithmes est que les agents apprennent de façon autonome. Lorsque l'algorithme est efficace, les performances peuvent s'approcher de celles obtenues par des humains sur les mêmes tâches, voir les surpasser sur certaines.

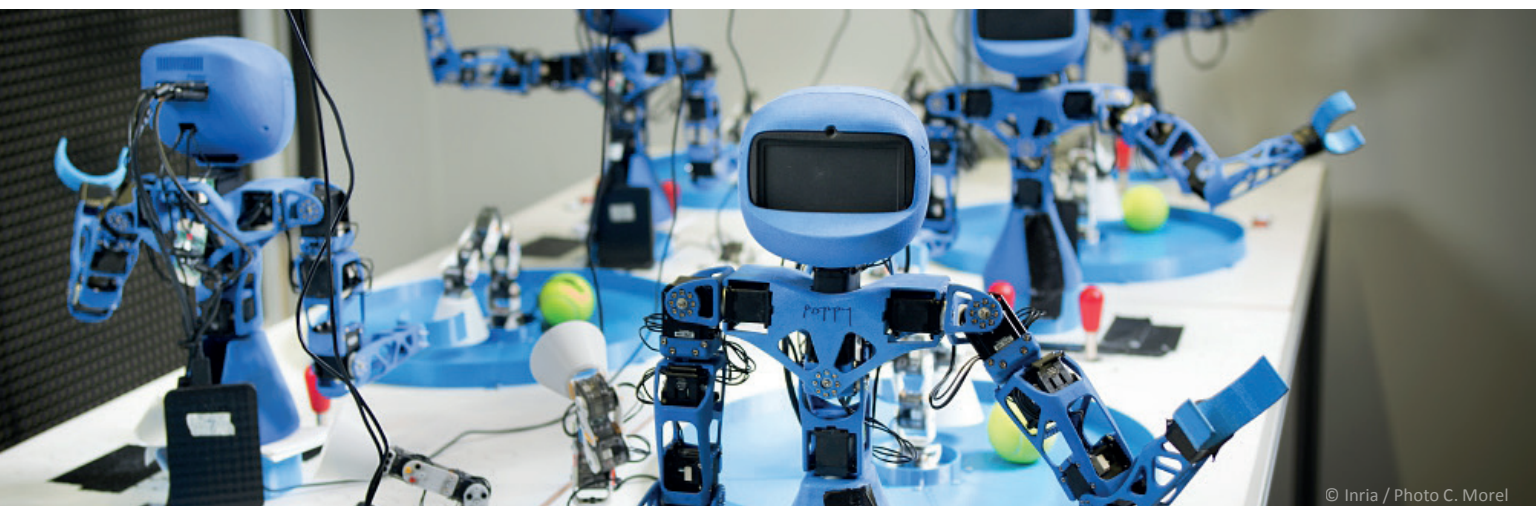
Exploration : l'agent essaie des actions variées pour augmenter sa connaissance de l'environnement afin de trouver la meilleure action à exécuter dans un état donné.

Exploitation : l'agent choisit l'action optimale selon ses connaissances actuelles pour maximiser sa récompense.

L'objectif est de trouver un compromis entre ces deux politiques afin de découvrir les meilleures actions sans sacrifier trop de récompenses.

AVANTAGES CONCURRENTIELS

- Apprentissage dans des espaces en grandes dimensions
- Capacité à résoudre des problèmes complexes
- Adaptabilité dans des environnements dynamiques
- Efficacité et autonomie des agents dans la prise de décision



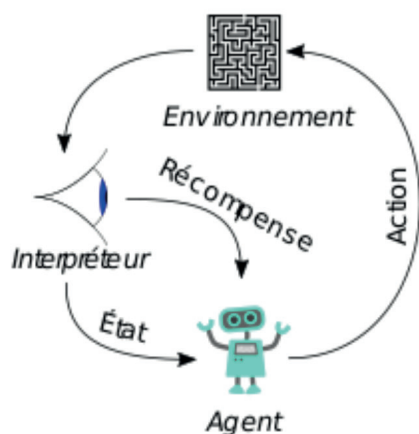


FICHE D'IDENTITÉ

- **Langage de programmation** : *Python, C, C++*
- **Bibliothèques** : *Gym, TensorFlow, PyTorch, Stable Baselines3 (SB3), CleanRL*
- **Équipe-projet** : *Flowers*
- **Compétences minimum requises** : *Gym* serait un plus

CAS D'USAGES

- **Robotique / voiture autonome** : déplacement autonome d'une machine
- **Santé** : stratégie d'intervention dans un cadre pandémique
- **Usine 4.0** : gestion complexe (pilotage des centrales électriques, optimisation de trajet, gestion de stock)
- **Finance** : gestion d'investissement en bourse
- **Jeux/jeux vidéo** : découverte de stratégie (jeu de Go), agent autonome dans les jeux vidéo, apprendre à jouer à un jeu vidéo en maximisant le score
- **Aéronautique/spatial/défense** : optimisation de trajectoires, coordination de systèmes interdépendants (flotte de drones, constellation de satellites, etc.).



FONCTIONNALITÉS GÉNÉRIQUES

- Optimise une politique d'action (c.-à-d. un comportement de l'agent) afin de maximiser une récompense sur le long terme dans un environnement dynamique
- Adapté lorsqu'il y a une interaction, constante, avec l'environnement pour la prise de décision en temps réel
- Environnement réel ou simulé pour extraire les cas
- Nécessite de la puissance de calcul
- Besoin d'un corpus de très grande taille, nécessite de nombreuses interactions avec l'environnement (plusieurs milliers, voire millions)

EXPERTS

- **Dan Dutartre** : ingénieur expérimentation et développement au sein du centre Inria de l'université de Bordeaux
- **Clément Moulin-Frier** : chargé de recherche chez Inria au sein de l'équipe-projet Flowers
- **Pierre-Yves Oudeyer** : directeur de recherche chez Inria au sein de l'équipe-projet Flowers

POUR ALLER PLUS LOIN :

- <https://www.deepmind.com/learning-resources/introduction-to-reinforcement-learning-with-david-silverhttps://spinningup.openai.com/en/latest/>
- https://www.youtube.com/playlist?list=PL_iWQOsE6TfX7MaC6C3HcdOf1g337dIC9
- <https://stable-baselines3.readthedocs.io/en/master/>
- <https://docs.cleanrl.dev/>