

Interaction Agent-Environnement dans l'Apprentissage par Renforcement (RL)

Introduction

L'interaction entre un **agent** et son **environnement** est au cœur de l'apprentissage par renforcement. Un agent est une entité capable de percevoir son environnement et de prendre des décisions. L'objectif de l'agent est de maximiser ses **récompenses** au cours du temps en choisissant les bonnes **actions** dans chaque situation (ou **état**).

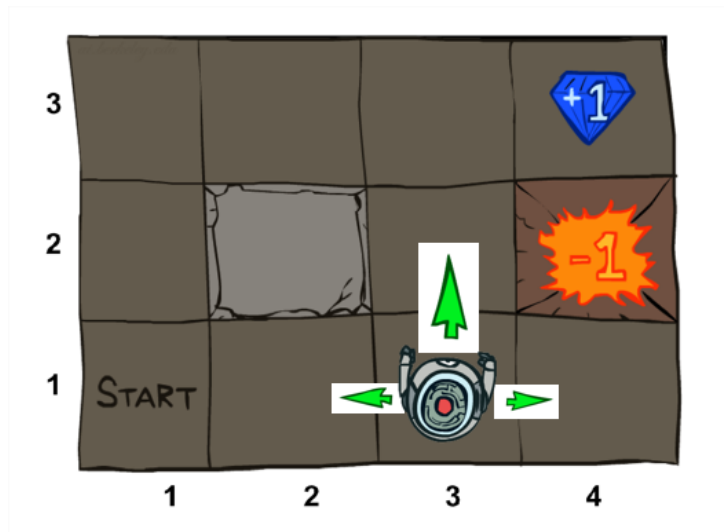


FIGURE 1 – Illustration de l'agent et de son environnement.

L'image ci-dessus représente un environnement avec différentes cases, certaines étant des obstacles, d'autres associées à des récompenses (+1) ou des pénalités (-1). L'agent (robot) commence à un point de départ et doit naviguer pour obtenir la meilleure récompense possible.

1 Notions de Base dans l'Interaction Agent-Environnement

1.1 États (States)

Un état représente une configuration ou une situation dans laquelle se trouve l'agent à un moment donné.

Exemple : Chaque case de l'environnement est un état. Si l'agent est en bas à gauche (1,1), cet état est unique.

1.2 Actions (Actions)

Une action est une décision prise par l'agent dans un état donné. **Exemple :** Se déplacer *gauche*, *droite*, *haut*, *bas*. Si l'agent est en (1,3), il peut aller vers la droite (1,4) ou vers le haut (2,3).

1.3 Récompenses (Rewards)

Les récompenses sont des retours de l'environnement après qu'une action a été prise.

Exemple : - Une case (3,4) avec une récompense de +1. - Une case (2,4) avec une pénalité de -1.

1.4 Transitions (Transitions)

Elles décrivent le passage d'un état à un autre en fonction de l'action choisie.

Exemple : Si l'agent est en (1,3) et choisit d'aller vers le haut, il passe à (2,3).

2 Exécution d'un Épisode

Un épisode représente une séquence d'interactions de l'agent avec l'environnement, de l'état initial à un état terminal.

2.1 Exemple de parcours de l'agent

Supposons que l'agent commence en (1,1) :

1. L'agent choisit d'aller à droite (2,1).
2. L'agent choisit encore droite (3,1).

3. L'agent choisit haut (3,2), proche d'une pénalité.
4. L'agent choisit droite (4,2) et reçoit une pénalité de -1.
5. L'agent choisit haut (4,3) et reçoit une récompense de +1.

2.2 Stratégie de l'agent

- L'agent apprend à éviter les cases avec pénalités et à atteindre la récompense de manière optimale.

3 Concept de Politique (Policy)

La **politique** est la stratégie adoptée par l'agent pour choisir ses actions à chaque état.

Exemple de politique optimale : - Éviter (4,2) qui contient une pénalité. - Aller directement à la case récompense (4,3).

4 Exploration et Exploitation

Exploration : Essayer différentes actions pour découvrir les meilleures stratégies. **Exploitation** : Appliquer une stratégie connue pour maximiser les récompenses.

Conclusion

L'apprentissage par renforcement repose sur l'interaction entre un agent et son environnement. Les notions d'**états**, **actions**, **récompenses** et **politique** sont essentielles et trouvent des applications en **robotique**, **gestion des systèmes intelligents** et **jeux vidéo**.

Exercice

- Simulez plusieurs parcours et analysez comment l'agent peut optimiser son itinéraire en évitant les pénalités.