

## **Exploration par échantillonnage uniforme en apprentissage par renforcement**

**Mots clés :** apprentissage par renforcement, gymnasium, exploration

**Encadrant :** Olivier Sigaud, ISIR, Sorbonne Université (Olivier.Sigaud at isir.upmc.fr)

**Nombre d'étudiants :** 2

**Résumé :** L'objectif de ce projet est de caractériser des environnements Gymnasium par la quantité d'exploration uniforme qu'ils requièrent.

### **Sujet développé :**

En apprentissage par renforcement, un agent apprend à contrôler un système de façon à maximiser un signal de récompense. Une difficulté centrale de ces méthodes est le compromis exploration – exploitation : à tout moment, l'agent doit choisir entre explorer pour découvrir des choses nouvelles, au risque de recevoir des récompenses faibles, ou bien exploiter les bonnes actions qu'il connaît déjà, au risque de ne jamais découvrir qu'il existe des actions encore meilleures.

Une méthode d'exploration idéale bien qu'irréaliste est l'échantillonnage uniforme : on met l'agent dans des états tirés aléatoirement et on lui fait réaliser des actions aléatoires. Cela permet de couvrir tous les couples (état, action) possibles sans se soucier du comportement déjà connu de l'agent.

Il existe un grand nombre d'environnements standards pour tester des algorithmes d'apprentissage par renforcement, regroupés dans la librairie Gymnasium. L'objectif de ce projet est de caractériser des environnements Gymnasium par la quantité optimale d'exploration uniforme qu'ils requièrent : dans les différents environnements, on entraînera des agents avec un mélange de données tirées du comportement courant de l'agent (pure exploitation) et de données tirées d'un échantillonnage uniforme (pure exploration). En faisant varier les quantités du mélange, on devrait obtenir une courbe de performance caractérisant chaque environnement. Par ailleurs, on pourra comparer l'exploration uniforme à une autre forme d'exploration basée sur un bruit gaussien ajouté à l'action courante : en faisant varier la variance de la gaussienne, on peut aussi passer de l'exploration pure (100 % de bruit) à 0 % (choix déterministe de l'action courante) et on fera la même étude.

Le codage s'effectuera en utilisant la librairie BBRL. Pour les étudiant.e.s qui seront choisi.e.s, les bénéfices de ce projet seront multiples :

- ce sera l'occasion de se familiariser avec les concepts et outils de l'apprentissage par renforcement, qui jouent un rôle significatif dans les progrès actuels de l'Intelligence Artificielle. Les étudiant.e.s verront en avance une partie du contenu de l'UE RL du M2 MIND.
- ce sera l'occasion de développer une bonne maîtrise de BBRL et d'apprendre une méthodologie de recherche rigoureuse.

Ce projet faisant appel à de nombreuses notions nouvelles, une formation initiale à l'apprentissage par renforcement et à des aspects méthodologiques sera assurée durant les premières semaines.

Une forte motivation pour l'apprentissage par renforcement, des bases en programmation python et une bonne maîtrise de github sont des prérequis.

### **Références**

- BBRL : <https://github.com/osigaud/BBRL>
- Gymnasium : <https://gymnasium.farama.org/>