APPRENTISSAGE PAR RENFORCEMENT SÉCURITAIRE : TRANSFERT SIM2REAL ET EXPLORATION SÚRE

A PREPRINT

Mouhamed Gando Diallo
IFT-7201 – Université Laval
mouhamed-gando.diallo.1@ulaval.ca

Renaud Djekornonde Raoudel
IFT-7201 – Université Laval
renaud.djekornonde-raoudel.1@ulaval.ca

29 avril 2025

Revue de la littérature

L'apprentissage par renforcement sécuritaire (Safe Reinforcement Learning, SRL) s'intéresse à l'apprentissage de politiques efficaces tout en évitant les comportements dangereux. Trois approches distinctes ont été analysées :

1. Vue d'ensemble du SRL. Gu et al. [**?**] classifient les dimensions du SRL à travers le cadre "2H3W" : *Why*, *What*, *How*, *When*, *Where*. Ils identifient les compromis entre exploration et sécurité, et mettent en évidence les défis algorithmiques et théoriques.

2. SRL dans un contexte industriel. Lu et al. [**?**] utilisent un algorithme RL avec contraintes de chance pour respecter des seuils critiques de sécurité dans une usine de traitement du minerai. Ils montrent que le RL peut être applicable aux environnements critiques, sous réserve de garanties de sécurité probabilistes.

3. Transfert Sim2Real avec mécanisme de Shielding. Hsu et al. [**?**] introduisent une architecture Sim-to-Lab-to-Real avec deux politiques : une pour la performance et l'autre pour bloquer les actions risquées. C'est cette approche qui est retenue pour notre projet.

Formulation du problème

Nous modélisons notre problème comme un processus de décision markovien $\mathcal{M} = (\mathcal{S}, \mathcal{A}, P, R, \gamma)$, avec :

- $\mathcal{S}$: espace des états
- $\mathcal{A}$: espace des actions continues
- P : dynamique (différente entre simulation et réalité)
- R : réwards avec pénalisation dés états dangereux
- γ : facteur d'actualisation

Nous cherchons à maximiser $\mathbb{E}_\pi[G]$ tout en contraignant :

$$\mathbb{P}_\pi(s_t \in \mathcal{S}_{\text{bad}}) \leq \delta \quad \forall t$$

Méthodes à l'étude

SAC – Soft Actor-Critic. Algorithme classique de RL pour actions continues, maximisant réward + entropie. Utilisé comme baseline via Stable-Baselines3.

SAC + Shielding. Wrapper qui bloque les actions menant à un angle ou une vitesse dangereuse. Deux versions : simple (seuil d'angle) et avancée (angle + vitesse angulaire).

Méthodologie expérimentale

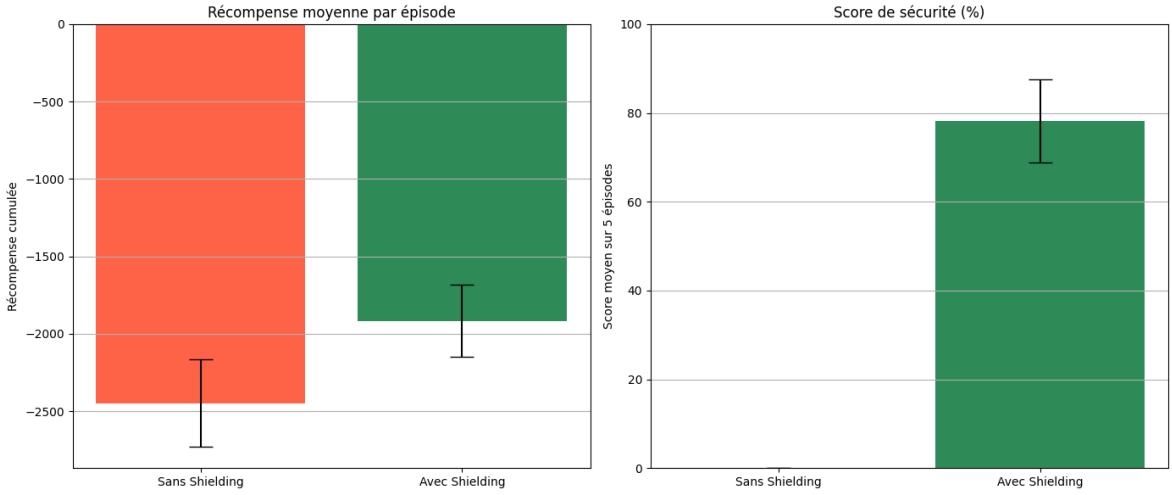
Entraînement. SAC est entraîné dans `Pendulum-v1` pendant 100k étapes, puis sauvegardé.

Transfert et Sécurité. La politique est testée dans `PendulumDangerous-v1`, un environnement avec $g = 15.0$, un moteur plus faible et un seuil d'angle dangereux à $160°$. Le Shielding bloque les actions si l'angle $> 130°$ ou $|\omega| > 8.0$.

Mésures. Pour chaque configuration :
— Réward moyen sur 5 épisodes
— Score de sécurité : proportion d'actions non bloquées

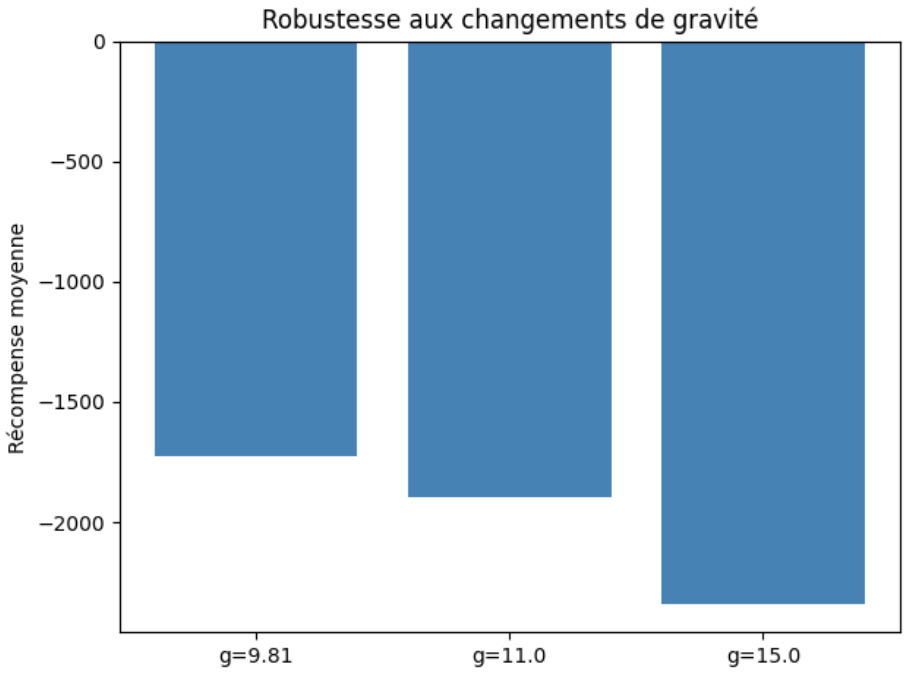
Résultats

Comparaison avec et sans Shielding.



Le Shielding réduit la fréquence des comportements dangereux (score ˜ 82%) au détriment de la performance (réward plus faible).

Transfert Sim2Real sous différentes gravités.



Quand la gravité augmente, la politique apprise en simulation dégrade fortement en performance. Cela met en évidence le besoin de robustesse et d'adaptation lors du transfert vers le réel.

Discussion

Les résultats confirment que les méthodes classiques de RL ne suffisent pas à garantir des comportements sécuritaires en situation de transfert. Le Shielding permet de limiter les risques de façon efficace mais au prix d'une baisse de performance. Ces observations soulignent l'intérêt d'une approche hybride entre performance et régulation sécuritaire dans des systèmes critiques.

Bibliographie

— Gu et al., *Safe Reinforcement Learning : A Survey*, 2022.
— Lu et al., *Safe RL with Chance Constraints in Gold Ore Processing*, 2023.
— Hsu et al., *Safe Continual Domain Adaptation after Sim2Real Transfer*, 2022.
— Haarnoja et al., *Soft Actor-Critic Algorithms*, 2018.

Projet GitHub

Le code complet du projet, incluant les expériences, les environnements et les scripts d'analyse, est disponible sur GitHub :

`https://github.com/gando537/Projet-RL-IFT-7201.git`