

Ordre du jour

TRAVAIL ACCOMPLI

- Finalisation du cahier des charges
- Ajouter Antmaze et dépendances
- Benchmark Lunar Lander mise à jour, compatibilité générique sur tous les environnements et expériences, pas encore de résultats, on devrait en avoir d'ici demain
- Parallélisation de l'entraînement, on passe de run qui durent 14h à 20m
- Fine-tuning avec Optuna paramétré par le nombre de trials pour chaque expérience
- Beaucoup de tentatives infructueuses de faire converger AFU Perrin sur Pendulum. En particulier, la fréquence d'updates à un impact étonnant sur la mesure de performance. Les précédents résultats utilisaient une mise à jour après chaque pas et on obtenait avec AFU Perrin une distribution à 2 gaussiennes, une proche de la politique optimale (-50.0) meilleure que la politique obtenue par DDPG et SAC et une distribution très mauvaise autour de la politique aléatoire. En changeant la fréquence de mise à jour (hyperparamètre), on obtient une politique très mauvaise autour de la politique aléatoire et on n'a plus de bons résultats ensuite. On a réalisé des histogrammes pour le visualiser. Pour résumer, en mettant à jour tout le temps, on a le potentiel d'une bonne politique (mais certainement pas en IQM) et lorsqu'on laisse Optuna gérer ce paramètre ou qu'on utilise celui du code de M. Perrin, on obtient de très mauvais résultats. On continue d'étudier AFU sur Pendulum et faire des études comparées avec Lunar Lander et Swimmer.

POINTS À DISCUTER

- Convergence d'AFU Perrin
- Discuter des premiers résultats de Off-On Policy qu'on obtiendra que d'ici demain. Nous nous excusons par avance de ne pas pouvoir les envoyer plus tôt
- Envoi d'ici à ce soir de la version finale du cahier des charges

OBJECTIFS

- A déterminer pendant la réunion
- Faire converger AFU Perrin