www.onera.fr

ONERA
THE FRENCH AEROSPACE LAB

PROPOSITION DE STAGE EN COURS D'ETUDES

Référence : **DTIS-2024-63**
(à rappeler dans toute correspondance)

Lieu : Toulouse

Département/Dir./Serv. : DTIS/SYD

Tél. : 05 62 25 26 00

Responsable(s) du stage : Filipo S. Perotto

Email. : filipo.perotto@onera.fr

DESCRIPTION DU STAGE

Thématique(s) : Intelligence Artificielle et Décision

Type de stage : ☒ Fin d'études bac+5 ☒ Master 2 ☐ Bac+2 à bac+4 ☐ Autres

Intitulé : PyRL - une library python open source pour federer des algorithmes d'apprentissage par renforcement deep et classiques en intégrant les méthodes dites Safe et Survival RL

Sujet : L'objet principal de ce stage est de continuer le developpement d'une librarie python open source souveraine pour federer les méthodes d'apprentissage par renforcement. Il s'agit de reimplementer certains algorithmes bien connus de la littérature, mais aussi de pouvoir integrer d'autres librairies python déjà existantes.

Scientifiquement, on vise developper des solutions pour le cas « survival reinforcement learning » (apprentissage par renforcement en survie), où un agent automatique doit choisir une séquence d'actions optimales pendant son interaction avec un environnement partiellement inconnu. Dans le domaine de la planification et de l'apprentissage, ces systèmes sont des Processus de Décision Markoviens (MDP). Dans l'opération de tels systèmes se pose la question de pouvoir apprendre en temps réel et sur place, tout en préservant son intégrité.

L'action de l'agent peut entraîner des récompenses positives et négatives, selon l'état du système et l'action choisie. L'agent dispose d'un budget initial qui change en fonction des récompenses reçues. L'objectif est donc de trouver un bon compromis entre exploration (i.e. agir pour apprendre de nouvelles choses), exploitation (i.e. agir de manière optimale en fonction de ce qui est déjà connu), et sécurité (garder à vue la gestion du budget), cherchant ainsi à apprendre à maximiser les récompenses au fil du temps, de façon efficiente, mais tout en minimisant le risque d'épuiser son budget.

Les algorithmes RL (apprentissage par renforcement) rencontrent un grand succès dans des environnements virtuels ou simulés, mais sont plus contraignants dans les problèmes concrets (systèmes cyber-physiques) où l'intégrité du système est un aspect critique, demandant de garanties de sécurité pendant le processus d'apprentissage.

Dans ce contexte, le stage veut aborder les défis scientifiques suivantes :

- La définition des stratégies d'apprentissage pour prendre en compte le risque de ruine.

Le plan de travail à suivre est le suivant :

- Etudier l'état de l'art scientifique et technique : RL, Deep RL, Safe RL, Survival RL.
- Implémenter des modèles choisis de la littérature dans un module python, faisant le lien avec d'autres librairies open-source
- Mettre en place des simulations comparatives à travers un exemple

Bibliographie :

[Perotto et al., 2019] Perotto, F., Bourgais, M., Silva, B., and Vercouter, L. (2019). Open problem: Risk of ruin in multiarmed bandits. In Proc. of COLT, pages 3194–3197.

[Perotto et al., 2021] Filipo Studzinski Perotto, Sattar Vakili, Pratik Gajane, Yaser Faghan, Mathieu Bourgais: Gambler Bandits and the Regret of Being Ruined. AAMAS 2021: 1664-1667.

[Majumdar and Pavone, 2020] Majumdar, A. and Pavone, M. (2020). How should a robot assess risk? towards an axiomatic theory of risk in robotics. Robotics Research, 10:75–84.

[Garcia and Fernandez, 2015] Garcia, J. and Fernandez, F.(2015). A comprehensive survey on safe reinforcement learning. JMLR, 16:1437–1480.

[Garcia and Shafie, 2020] Garcia, J. and Shafie, D. (2020). Teaching a humanoid robot to walk faster through safe reinforcement learning. Engineering Applications of Artif. Intel., 88.

[Efroni et al., 2020] Efroni, Y., Mannor, S., and Pirotta, M. (2020). Exploration-exploitation in constrained mdps. ArXiv, abs/2003.02189.

[Caramanis et al., 2014] Caramanis, C., Dimitrov, N., and Morton, D. (2014). Efficient algorithms for budget constrained markov decision processes. IEEE Trans. Automat. Contr., 59(10):2813–2817.

[Carpin et al., 2016] Carpin, S., Chow, Y., and Pavone, M. (2016). Risk aversion in finite markov decision processes using total cost criteria and average value at risk. In Proc.of ICRA, pages 335–342. IEEE.

Est-il possible d'envisager un travail en binôme ? **Non**

Méthodes à mettre en oeuvre :

☐ Recherche théorique	☒ Travail de synthèse
☒ Recherche appliquée	☒ Travail de documentation
☒ Recherche expérimentale	☒ Participation à une réalisation

Possibilité de prolongation en thèse : **Oui**

Durée du stage : Minimum : 5 Maximum : 6

Période souhaitée : Février-Juillet

PROFIL DU STAGIAIRE

Connaissances et niveau requis :	Ecoles ou établissements souhaités :
Intelligence Artificielle, Apprentissage par Renforcement, Programmation Python, niveau M2	Université de Toulouse, ISAE, ENSEEIHT