

Le projet doit être réalisé en **équipe de deux** étudiants-es.
Remise analyse préliminaire : Lundi 30 mars à 23h59.
Remise rapport final : Vendredi 1er mai à 23h59.

1 Mise en contexte

Vous travaillez dans une firme de consultants en intelligence artificielle. Un client s'intéresse à l'utilisation de l'apprentissage par renforcement (RL) dans des environnements non-déterministes, où des situations dangereuses doivent être évitées. Plus précisément, votre client s'intéresse à des applications dans lesquelles les situations dangereuses doivent être évitées le plus possible *durant l'apprentissage*, puisque celles-ci peuvent coûter très cher. Il vous demande donc de réaliser une étude montrant les limitations des stratégies de RL sous ces contraintes et d'évaluer des approches proposées dans la littérature pour aborder ce problème.

2 Travail à réaliser

L'objectif du projet consiste à étudier l'impact des dynamiques non-déterministes dans un environnement sur la capacité d'évitement des situations dangereuses lors de l'apprentissage d'une politique via RL. Le projet doit être réalisé en Python sur l'environnement **Frozen Laze (v1)** de la librairie **Gymnasium** [Towers et al., 2024]. Dans cet environnement, un joueur doit traverser un lac gelé comportant des trous (situations dangereuses) qui doivent être évités. La surface glissante du lac limite la prévisibilité des déplacements (environnement non-déterministe). On cherche une stratégie d'apprentissage permettant au joueur d'éviter de tomber dans les trous durant l'apprentissage de sa politique.

Votre projet doit être hébergé sur **GitHub**. Nous vous communiquerons les noms d'utilisateurs à ajouter au projet en vue de la correction. Assurez-vous de bien utiliser la commande **commit** puisque l'historique des modifications sera utilisé pour valider vos contributions¹ au projet et contrôler le plagiat.

Le travail sera effectué en deux étapes : 1) l'analyse préliminaire; 2) le rapport final.

2.1 Analyse préliminaire

Dans un premier temps, votre client vous demande de produire une analyse préliminaire visant à démontrer une prise en main de l'environnement Frozen Lake via la librairie Gymnasium, décrire votre étude à réaliser et présenter les résultats d'expériences préliminaires. Votre document doit suivre la structure suivante :

- Cas d'étude
- Stratégies considérées
- Méthodologie expérimentale
- Résultats
- Bibliographie

Votre document peut contenir **jusqu'à 4 pages** (excluant la bibliographie) et doit être rédigé en LaTeX dans le style suivant : <https://github.com/kourgeorge/arxiv-style>. Ne pas inclure le résumé (*abstract*) ni les mots-clés (*keywords*). Vous devez remettre l'analyse préliminaire en format PDF. Votre document doit contenir l'URL du projet **GitHub** contenant le code source (Python) documenté pour reproduire vos expériences.

Cas d'étude L'environnement Frozen Lake est très configurable. Il est de votre responsabilité de bien comprendre l'environnement pour proposer une étude pertinente à réaliser qui permettra d'évaluer l'impact d'une dynamique d'environnement non-déterministe sur l'évitement de situations dangereuses. À cet effet,

¹Pénalité possible jusqu'à 50% pour non-contribution.

présentez les composantes de l'environnement en expliquant comment celles-ci peuvent être manipulées pour ajuster la difficulté de l'environnement : degré de stochasticité dans la dynamique; nombre de situations dangereuses; positionnement des situations dangereuses. Finalement, présentez **trois cas d'étude** que vous souhaitez considérer et expliquez concrètement comment configurer votre environnement pour les obtenir. Décrivez les niveaux de difficultés/défis capturés par vos cas d'étude pour motiver leur pertinence.

Stratégies considérées Dans le cadre du projet, vous devez comparer différentes stratégies de RL. À l'étape de l'analyse préliminaire, identifiez **deux stratégies** (p.ex., DQN [Mnih et al., 2015], PPO [Schulman et al., 2017]) qui serviront de *baselines* pour l'étude. Nous faisons référence ici à des stratégies qui n'ont pas été conçues explicitement avec des mécanismes pour gérer les défis retrouvés dans cet environnement. Motivez le choix de ces stratégies et présentez leur fonctionnement.

Méthodologie expérimentale Vous devez réaliser des expériences avec les stratégies considérées sur les différents cas d'études établis. Pour chaque stratégie considérée, vous fournissez le détail des hyper-paramètres utilisés et comment ils ont été déterminés. Vous pouvez utiliser des librairies existantes qui fournissent une implémentation des stratégies, comme **Stable-Baselines3** [Raffin et al., 2021]. Indiquez simplement quelles sont ces librairies et comment vous configurez les différentes stratégies. Décrivez également tous les paramètres expérimentaux nécessaires à la reproduction de votre étude. Cela inclut notamment la durée des épisodes, le nombre de pas de temps pour l'apprentissage et le nombre de répétitions (pour obtenir des intervalles de confiance sur la performance). Finalement, présentez les différents indicateurs de performance que vous utilisez pour quantifier l'atteinte des objectifs de l'étude par les stratégies considérées. Attention, l'objectif ne se limite pas à maximiser les récompenses sur la politique finale!

Résultats Présentez les résultats obtenus en suivant la méthodologie décrite précédemment. Vos résultats doivent inclure des tableaux et/ou figures qui permettent de suivre les indicateurs de **performance durant l'apprentissage** sur les différents cas d'étude considérés et visualiser la **différence entre les politiques** apprises par les deux stratégies. Dans votre présentation des résultats, indiquez les faits saillants à remarquer sur chaque tableau/figure. Cette étape est cruciale puisqu'elle informera votre sélection de stratégies subséquentes pour améliorer la performance (en vue du rapport final).

Bibliographie Vous devez fournir les références bibliographiques appropriées lorsque vous référez à des éléments de la littérature qui ne vous appartiennent pas. Cela inclut par exemple les environnements, les librairies et les algorithmes,stratégies. Assurez-vous d'uniformiser le style de vos entrées dans la bibliographie. Évitez également d'inclure des informations inutiles (p.ex., DOI, ISSN, URL).

2.2 Rapport final

Suite à une validation de l'analyse préliminaire par le client, vous devez compléter votre étude pour y inclure une stratégie de RL disposant de mécanismes particuliers pour aborder les défis du problème considéré. Votre rapport final doit suivre la structure suivante :

- Cas d'étude (mise à jour)
- Stratégies considérées (mise à jour)
- Méthodologie expérimentale (mise à jour)
- Résultats (mise à jour)
- Discussion
- Bibliographie (mise à jour)

Votre document peut contenir **jusqu'à 6 pages** (excluant la bibliographie) et doit être rédigé en LaTeX dans le style suivant : <https://github.com/kourgeorge/arxiv-style>. Ne pas inclure le résumé (*abstract*)

ni les mots-clés (*keywords*). Vous devez remettre le rapport final en format PDF. Votre rapport doit contenir l'URL du projet [GitHub](#) contenant le code source (Python) documenté pour reproduire vos expériences.

Cas d'étude Aucun nouveau matériel n'est attendu à cette étape. Appliquez simplement les modifications suivant la correction de l'analyse préliminaire le cas échéant.

Stratégies considérées En vue du rapport final, fouillez la littérature scientifique et **identifier trois articles** publiés (journaux ou conférences scientifiques) présentant chacun une méthode adaptée au problème visé par l'étude. Décrivez brièvement (3-5 phrases) chaque stratégie. Sélectionnée ensuite celle que vous jugez la plus prometteuse pour améliorer la performance des stratégies *baselines* dans le problème considéré. Justifiez le choix de la stratégie sélectionnée et présentez son fonctionnement. Ajoutez ce nouveau matériel à la suite des stratégies *baselines* mises à jour suivant la correction de l'analyse préliminaire.

Méthodologie expérimentale Vous devez réaliser des expériences avec les stratégies considérées sur les différents cas d'études établis. Pour chaque stratégie considérée, fournez le détail des hyper-paramètres utilisés et comment ils ont été déterminés. Vous pouvez utiliser des implémentations existantes des stratégies disponibles via des librairies ou sur GitHub. Indiquez simplement les sources de vos implémentations et comment vous les utilisez. Si des erreurs méthodologiques ont été relevées à l'étape du rapport préliminaire, assurez-vous de les corriger, en mettant à jour la description de la méthodologie, et de refaire les expériences avec les stratégies *baselines*. Cela inclut notamment les indicateurs de performance

Résultats Présentez les résultats obtenus en suivant la méthodologie décrite précédemment sur l'ensemble des stratégies considérées. Vos résultats doivent inclure des tableaux et/ou figures qui permettent de suivre les indicateurs de **performance durant l'apprentissage** sur les différents cas d'étude considérés et visualiser la **différence entre les politiques** apprises par les différentes stratégies. Dans votre présentation des résultats, indiquez les faits saillants à remarquer sur chaque tableau/figure.

Discussion Discutez des résultats obtenus en fonction de ce qui était attendu d'après la littérature. Expliquez les résultats obtenus (ne vous contentez pas de mentionner que l'approche A est meilleure que l'approche B). N'hésitez pas à référer à la littérature pour supporter vos hypothèses. Présentez votre réflexion sur la pertinence des différentes approches, en considérant la complexité algorithmique, les défis d'implémentation, ainsi que le coût en calcul. Présentez vos conclusions générales relatives aux difficultés capturées par votre étude : Quel serait votre message général au client qui a commandé cette étude?

Bibliographie Mettre à jour suivant les mêmes instructions que pour l'analyse préliminaire.

References

- Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015.
- Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, 22(268):1–8, 2021.
- John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.
- Mark Towers, Ariel Kwiatkowski, Jordan Terry, John U Balis, Gianluca De Cola, Tristan Deleu, Manuel Goulão, Andreas Kallinteris, Markus Krimmel, Arjun KG, et al. Gymnasium: A standard interface for reinforcement learning environments. *arXiv preprint arXiv:2407.17032*, 2024.