

Sujet de thèse 2024-2027 à l'INSA de Rouen Normandie
**Optimisation de l'apprentissage de politiques de décision
pour une équipe de robots sociaux**

Contexte

L'apprentissage par renforcement (RL) est une technique d'apprentissage non supervisé dans laquelle un agent apprend une politique d'action par essais et erreurs, en obtenant une récompense de l'environnement pour les actions entreprises. Le domaine du problème, c'est-à-dire l'espace d'entrée, est souvent modélisé à l'aide d'un ensemble fini d'états, et l'objectif de l'agent est, à chaque étape, de choisir l'action qui maximise la récompense à long terme pour un état donné. Pour modéliser un problème à l'aide d'un processus RL, la discrétisation de l'espace des états, des actions et des récompenses doit être spécifiée. La méthode la plus courante est de discrétiser manuellement l'espace des états et des actions, en utilisant des connaissances expertes du domaine. Cela n'est cependant pas facilement réalisable dans des situations où l'espace d'états et l'espace d'actions doivent être dérivés automatiquement à partir des entrées/sorties du système (perception et contrôle bruts), ou quand ils doivent s'adapter au fur et à mesure des interactions entre l'agent et son environnement, potentiellement composé d'autres agents.

Les états représentent alors une vision imparfaite/floue de l'état du système, biaisée par une connaissance incomplète ou imparfaite de la réalité (POM) ; cette représentation peut elle-même être distribuée sur plusieurs entités, par exemple lors de l'exploration d'un espace inconnu par une flotte d'agents. Lorsque la politique de prise de décision repose sur un mécanisme par renforcement (POM-DP), **la difficulté majeure demeure la taille de l'espace à explorer.**

Objectifs

L'objectif principal de cette thèse est d'optimiser l'apprentissage de la politique de prise de décision en diminuant l'espace d'exploration et en répartissant cette exploration auprès des différents agents.

La première des pistes envisagées est d'utiliser des algorithmes d'apprentissage non supervisé afin de (S1) **diminuer l'espace d'exploration** par regroupement d'états représentant des contextes similaires (axe 1 : stratégie guidée par simplification de l'espace des états). La littérature propose principalement des approches basées sur fonctions d'approximation pour adapter l'espace des états en RL (Samejima 1999), mais la génération de ces fonctions rend l'abstraction peu interprétable pour les systèmes qui les utilisent. Une autre solution est d'utiliser des algorithmes d'apprentissage non-supervisés (type clustering) afin de générer dynamiquement de nouvelles partitions de la perception des agents (Viera 2013, Guériaux 2019). La problématique qui reste à explorer est d'améliorer la partie génération tout en permettant à ce type de représentation plus abstraite d'être transférable entre différents agents, cela afin de faciliter l'apprentissage dans des contextes similaires explorés en parallèle par plusieurs systèmes.

La seconde piste consiste en la (S2) **génération de plans de plus haut niveau**, soit en exploitant une hiérarchie d'actions, soit en identifiant des sous-séquences similaires (axe 2 : stratégie guidée par simplification de l'espace des actions). La littérature recense plusieurs types d'approches qui pourraient être utilisées ici, sous le champ de recherche de l'« apprentissage hiérarchique » (Barto 2003). Parmi les plus pertinentes, on retrouve

l'agrégation d'actions sous la forme d'options (Sutton 1999) ou encore de compétences (Thrun 1994). Dans ce travail de thèse, nous souhaitons proposer des méthodes d'apprentissage (ou de génération) de ces abstractions tout en permettant le partage de ces représentations entre différents agents, dans la veine des travaux connexes sur les techniques (Konidaris 2007) et la partie applicative (Saleh 2020).

Ces deux stratégies de réduction de l'espace d'exploration (S1 et S2), associées à une répartition intelligente entre différents agents (type apprentissage par transfert (Taylor 2019)), devraient permettre d'accélérer leur processus d'apprentissage et de limiter et/ou mitiger la problématique du « curse of dimensionality » (Barto 2003) qui anime la communauté RL depuis plusieurs années.

D'un point de vue applicatif, il s'agira d'exploiter cette approche théorique et de la valider lorsque plusieurs robots sociaux en interaction avec des utilisateurs doivent apprendre à réaliser de la meilleure manière possible une tâche collective d'assistance (par exemple, dans le cadre de la Robocup@Home, prendre des commandes de cocktails auprès d'utilisateurs, préparer ces cocktails et les servir ; et afin de démontrer la généralisation de ce type d'approche à d'autres contextes : l'exploration collaborative par des robots mobiles). Du point de vue de l'acceptabilité, la formalisation de la politique de prise de décision par des plans (abstraites ou non) permettrait de faciliter l'explicabilité des prises de décision (politique) auprès des utilisateurs, ce qui a été identifié comme une piste essentielle pour les approches à base de RL (Milani 2023, Zelvelde 2021).

Contributions attendues et valorisation

Les prototypes développés : d'abord en simulation, puis en reprenant des briques existantes sur les robots sociaux et/ou mobiles ; serviront à la fois de démonstrateurs (journées portes ouverte, fête de la science, accueil d'industriels ...), mais permettront aussi la poursuite d'autres travaux de recherche. Par exemple, une perspective évidente serait la transposition des modèles développés dans cette thèse sur une flotte de robots mobiles modélisant des configurations/problématiques rencontrées par le véhicule autonome, où une collaboration entre les véhicules via l'élaboration de plans/politiques explicables, simplifiées et/ou plus abstraites pourrait faciliter la prise de décision dans des contextes particuliers (ex : scénarios limites qui mettent en défaut la perception).

Références

- (Arnold 2021) Arnold, T., Kasenberg, D., Scheutz, M. (2021). Explaining in time: Meeting interactive standards of explanation for robotic systems. *ACM Transactions on Human-Robot Interaction (THRI)*, 10(3), pp.1-23.
- (Barto 2003) Barto, A. G., & Mahadevan, S. (2003). Recent advances in hierarchical reinforcement learning. *Discrete event dynamic systems*, 13(1-2), 41-77.
- (Guérliau 2019) Guérliau, M., Cardozo, N., & Dusparic, I. (2019). Constructivist approach to state space adaptation in reinforcement learning. In *2019 IEEE 13th International Conference on Self-Adaptive and Self-Organizing Systems (SASO)* (pp. 52-61). IEEE.
- (Malik 2021) Malik U, Barange M, Saunier J, Pauchet A: A novel focus encoding scheme for addressee detection in multiparty interaction using machine learning algorithms. *J. Multimodal User Interfaces* 15(2): 1-14 (2021).
- (Milani 2023) Milani, S., Topin, N., Veloso, M., & Fang, F. (2023). Explainable reinforcement learning: A survey and comparative review. *ACM Computing Surveys*.
- (Konidaris 2007) Konidaris, G. D., & Barto, A. G. (2007, January). Building Portable Options: Skill Transfer in Reinforcement Learning. *IJCAI (Vol. 7, pp. 895-900)*.
- (Orr 2023) Orr, J., Ayan D. (2023). Multi-agent deep reinforcement learning for multi-robot applications: A survey. *Sensors (vol 23.7, pp. 3625)*.

- (Rasendrasoa 2022) Rasendrasoa S, Pauchet A, Saunier J, Adam S: Real-Time Multimodal Emotion Recognition in Conversation for Multi-Party Interactions. ICMI 2022: 395-403.
- (Saleh 2020) Saleh, A., Jaques, N., Ghandeharioun, A., Shen, J., & Picard, R. (2020). Hierarchical reinforcement learning for open-domain dialog. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 05, pp. 8741-8748).
- (Samejima 1999) Samejima, K., & Omori, T. (1999). Adaptive internal state space construction method for reinforcement learning of a real-world agent. *Neural Networks*, 12(7-8), 1143-1155.
- (Sutton 1999) Sutton, R. S., Precup, D., & Singh, S. (1999). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112(1-2), 181-211.
- (Taylor 2009) Taylor, M. E., & Stone, P. (2009). Transfer learning for reinforcement learning domains: A survey. *Journal of Machine Learning Research*, 10(7).
- (Taylor 2019) Taylor, A., Dusparic, I., Guériau, M., & Clarke, S. (2019, July). Parallel transfer learning in multi-agent systems: What, when and how to transfer?. In 2019 International Joint Conference on Neural Networks (IJCNN) IEEE.
- (Thrun 1994) Thrun, S., & Schwartz, A. (1994). Finding structure in reinforcement learning. *Advances in neural information processing systems*, 7.
- (Viera 2013) Vieira, D. C. D. L., Adeodato, P. J. L., & Goncalves, P. M. (2013). A temporal difference gng-based approach for the state space quantization in reinforcement learning environments. In 2013 IEEE 25th International Conference on Tools with Artificial Intelligence (pp. 561-568). IEEE.
- (Zelvelde 2021) Zelvelde, A. E., Westberg, M., & Främling, K. (2021). Assessing explainability in reinforcement learning. In International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems (pp. 223-240). Springer International Publishing.

Mots-clefs

Abstraction de représentation, apprentissage par renforcement, apprentissage explicable, apprentissage hiérarchique, robotique sociale

Profil recherché

- Master en informatique (ou en robotique), avec une spécialisation ou un intérêt pour les techniques basées sur l'IA distribuée et/ou l'apprentissage par renforcement ;
- Solides compétences en communication écrite et orale en anglais et/ou en français.

Des connaissances et/ou une expérience dans les domaines suivants seraient grandement appréciées :

- Systèmes multi-agents ;
- Techniques d'apprentissage (par renforcement, profond ou non) ;
- Environnements robotiques.

Encadrement

Alexandre Pauchet (Maître de conférences, HDR) et Maxime Guériau (Maître de conférences) à l'[INSA de Rouen Normandie](#), membres des équipes Multi-Agent, Interaction, Décision ([MIND](#)) et Systèmes de Transport Intelligents ([STI](#)) au [LITIS](#) (*Laboratoire d'Informatique, de Traitement de l'Information et des Systèmes*).

A propos du LITIS

La recherche menée au laboratoire LITIS couvre trois grands domaines : l'accès à l'information, le traitement de l'information biomédicale et l'intelligence ambiante avec des applications dans la santé, du véhicule intelligent et les territoires intelligents. L'expertise des membres du LITIS est reconnue internationalement et comprend : l'apprentissage automatique, les systèmes multi-agents, les véhicules intelligents. Le LITIS est un laboratoire (EA 4108) de l'Université de Rouen Normandie, de l'Université du Havre Normandie et de l'INSA Rouen Normandie. Il est membre de l'école doctorale MIIS et du réseau normand "Digital Normandy". Le LITIS est partenaire de la Fédération de recherche CNRS Normastic.

Le candidat aura accès à différentes plates-formes expérimentales pour mener ses travaux :

- Une plateforme robotique comprenant différents véhicules robotisés équipés de capteurs de perception de pointe ;
- Un centre informatique intensif (CRIANN : Centre Régional Informatique et d'Applications Numériques de Normandie).

Financement

Cette thèse sera financée pendant 3 ans par l'INSA de Rouen Normandie (50%) et la région Normandie (50%). Le doctorant retenu sera inscrit à l'[école doctorale MIIS](#) (ED 590), qui propose des formations obligatoires ainsi qu'un accompagnement pendant la thèse.

Processus de candidature

Les candidatures doivent comprendre :

- un CV complet et à jour, avec une liste des publications éventuelles ;
- une lettre de motivation ;
- les informations de contact de 1-2 personnes références, si disponible ;
- un relevé de notes du Master ;

et être envoyées par email avec l'objet « candidature thèse Appodéroso » à :

alexandre.pauchet@insa-rouen.fr

maxime.gueriau@insa-rouen.fr

au plus tard le vendredi 24 mai 2024.