Sujet de thèse: Exploration et construction de représentations en environnements ouverts.

Depuis une vingtaine d'années, les architectures de calcul parallèle sont devenues nettement plus efficaces, plus facilement utilisables et de moins en moins chères. Sur la même période, d'importants progrès théoriques et expérimentaux en apprentissage machine ont permis d'exploiter au mieux ces nouveaux moyens de calculs. Progressivement, dans de nombreux domaines, les performances maximales sur des problèmes de référence ont été établies par des algorithmes d'apprentissage automatique. C'est le cas notamment en reconnaissance d'image, en traitement du langage naturel, et en apprentissage par renforcement, un sous-domaine de l'intelligence artificielle pouvant avoir une influence importante sur la robotique. Les algorithmes d'apprentissage par renforcement permettent de découvrir de façon automatique, par essais et erreurs, des comportements atteignant un objectif donné sous la forme d'une récompense à maximiser. Ils peuvent donc donner aux robots la capacité à s'adapter à différentes situations, mais pour ce faire, ils nécessitent de connaître une représentation adaptée du problème avec des états, des actions et un signal de récompense appropriés. Cependant, malgré des progrès très significatifs, ces approches n'ont pour l'instant pas abouti à un déploiement à grande échelle de techniques d'apprentissage sur des agents ou robots évoluant dans des environnements ouverts, dans lesquels les tâches sont rarement précisément définies.

Pour permettre à des agents d'évoluer et de s'adapter correctement dans des environnements ouverts, il semble nécessaire de repenser les objectifs de l'apprentissage par renforcement. Les agents sont habituellement évalués sur des tâches identiques à celles sur lesquelles ils ont été entraînés, ou sur des tâches provenant d'une distribution de tâches correspondant à celle de la phase d'entraînement. Dans le cas d'un environnement totalement ouvert, même l'hypothèse d'une distribution de tâches connues est trop forte, car l'agent peut être amené à rencontrer des tâches radicalement différentes, et doit en permanence explorer et s'adapter. **L'objectif de cette thèse est de faire progresser les techniques d'apprentissage en environnement ouvert.** La question principale n'est pas d'augmenter la performances sur des tâches précises, mais plutôt de parvenir à une adaptation permanente en présence de conditions changeantes et incertaines. Spécifiquement, cette thèse propose d'aborder simultanément deux problèmes importants: **l'exploration et la construction de représentations.**

Exploration et construction de représentations

Les capteurs et effecteurs du robot permettent d'envisager de nombreux espaces d'états et d'action différents. Leur définition est un compromis entre l'éventail des possibilités qu'ils proposent et la difficulté à explorer efficacement des espaces de grande dimension. Un apprentissage "ouvert" doit donc pouvoir explorer les comportements possibles à partir de représentations au plus près des perceptions générées par les capteurs du robot et des commandes attendues par ses moteurs. Ce type d'apprentissage bout à bout (end-to-end) soulève de nombreuses questions sur les deux dimensions complémentaires que cette thèse va considérer: l'exploration et la construction de représentations.

Des représentations compactes et structurées permettent de réduire la complexité de l'apprentissage par renforcement, mais si ces représentations ignorent des informations essentielles, elles peuvent au contraire rendre tout apprentissage impossible. Pour distinguer les informations les plus pertinentes et construire une représentation la plus efficace possible, l'agent doit explorer une grande diversité de situations, c'est pourquoi la construction de représentations et l'exploration sont deux problèmes intimement liés, bien qu'ils soient habituellement traités séparément. Les représentations apprises durant l'exploration devraient non seulement permettre d'accélérer l'apprentissage, mais aussi de mieux explorer. Ce thème, que nous pouvons formuler **"apprendre à explorer"**, sera l'un des axes principaux de la thèse. Le second axe principal aura trait à la **généralisation**. En effet, en environnement ouvert, on attend également des représentations qu'elles permettent à l'agent de facilement transférer ses aptitudes lorsque le signal de récompense est modifié.

Les travaux se baseront sur des résultats récents de l'équipe d'accueil [9], qui ont démontré que la gestion simultanée de l'exploration et de la construction de représentation permet d'aboutir à des représentations qui facilitent par la suite la résolution de tâches d'apprentissage par renforcement.

1) Représentations et généralisation

La question principale de cet axe de recherche est la suivante: **comment de bonnes représentations peuvent-elles favoriser la généralisation des aptitudes acquises ?**

Les représentations peuvent concerner l'espace des états ou l'espace des actions.

L'état du robot est l'information utile pour qu'il puisse prendre la décision appropriée ou contrôler ses mouvements. Les observations issues des capteurs du robot ne suffisent pas en général, car les capteurs peuvent être sujets à du bruit, des occlusions ou de l'aliasing perceptif (plusieurs conditions différentes donnant lieu aux mêmes obervations). Il est donc nécessaire de construire un état qui contienne une information plus fiable et plus stable. La construction automatique d'espaces d'état s'appuie sur différents principes, par exemple la reconstruction des observations après être passé par un goulot d'étranglement (auto-encodeurs), l'apprentissage d'un modèle direct ou inverse, ou encore l'utilisation de pro-

priétés "physiques" des états [6]. Ces différentes approches construisent des états répondant à des besoins variés prenant en compte, ou pas, la dynamique du robot et ses capacités d'action. Elles sont testées principalement dans des environnements simplifiés et leur transfert à des robots réels reste un sujet ouvert. Les travaux de thèse s'appuieront sur la littérature de ce domaine et aborderont ce sujet sous un angle original, celui d'une démarche itérative, dans la continuité de l'approche proposée pour l'apprentissage ouvert dans l'équipe d'accueil [3, 2]. Une approche possible pourra être de changer les fonctions de coût pilotant la construction de l'espace d'état au cours du processus en commençant par des motivations visant tout d'abord à construire un état lié aux capacités d'action du robot, puis à ses capacités de prédiction et enfin à la capacité à reconstruire les perceptions. Ce *curriculum* vise à augmenter progressivement les capacités de prédiction des modèles appris en fonction de la quantité de données acquises pour tendre vers les modèles les plus généraux possibles, modèles qui permettront de généraliser les aptitudes acquises.

L'espace des actions définit ce que le robot est capable de faire. L'apprentissage de cet espace est fortement lié à l'apprentissage de l'espace d'état: les actions doivent permettre de se déplacer efficacement dans l'espace des états pour que le robot puisse atteindre un objectif fixé. Cette question sera donc abordée en même temps que celle de la construction d'espaces d'états.

2) Apprendre à explorer

La question principale de cet axe de recherche est la suivante: **dans un environnement ouvert, comment apprendre à progressivement améliorer l'efficacité de l'exploration ?**

Les stratégies d'exploration les plus simples s'appuient sur des mouvements aléatoires ou sur un gradient de récompense. Dans des cas réalistes, les mouvements aléatoires ne génèrent que peu de mouvements intéressants [8] et une récompense est définie sur la base d'un espace d'état et d'un espace d'action. Lorsque ces espaces sont en cours de construction, la définition de la récompense est donc délicate. De plus, les récompenses sont souvent rares en robotique. Dans le cas de l'apprentissage de la saisie d'objets, par exemple, très peu de mouvements d'un bras réussiront à attraper l'objet et donc à générer de la récompense. Écrire une récompense dense, susceptible de guider un processus d'exploration, nécessite de disposer d'une connaissance fine de la tâche que l'on souhaite réaliser, ce qui n'est pas le cas en apprentissage ouvert. Des méthodes d'explorations basées sur l'algorithme de recherche de nouveauté [5] et les algorithmes de qualité et diversité [12, 1] permettent de réaliser une telle exploration lorsque l'on ne dispose pas de récompense dense [4, 10], il est même possible de les combiner à la recherche d'un espace comportemental [11].

L'objectif de cette thèse est de poursuivre les travaux commencés dans l'équipe d'accueil et d'utiliser des méthodes de qualité et diversité dans chacune des itérations de la méthode d'apprentissage d'espace d'état. Ces méthodes peuvent en effet générer des données variées facilitant l'apprentissage de modèle [7].

Déroulement de la thèse

1ère année

Le candidat va réaliser un stage de master sur la thématique de la thèse. Il portera sur une première étape consistant à passer à l'échelle les méthodes développées pendant la thèse d'Astrid Merckling [9] en intégrant les méthodes d'exploration développées pendant la thèse de Giuseppe Paolo [11] pour, en même temps, explorer, construire un espace d'état et un modèle prédictif pour un système robotique incluant des interactions rares. La première année de la thèse sera consacrée à la poursuite de ce travail afin de le publier dans une conférence du domaine (Corl, ICLR, NeurIPS, GECCO, IEEE-IROS ou IEEE-ICRA) ou dans un journal international (IEEE-TRO, JMLR, Robotics and Autonomous systems ou encore Frontiers in Robotics and AI).

En parallèle, le candidat réalisera une étude bibliographique poussée du sujet et montera en compétence au travers de l'implémentation de méthodes de l'état de l'art. Il réalisera également une analyse du problème et choisira le cadre expérimental permettant de démontrer l'intérêt des algorithmes qui seront développés dans la suite de la thèse. Ces travaux, dans la continuité des travaux réalisés pendant le stage de master pourront donner lieu à une autre publication en fin de première année.

Selon l'avancement, la préparation d'un article de synthèse des méthodes de l'état de l'art sera considérée.

2ème année

La deuxième année sera consacrée à la consolidation des travaux réalisés. Selon les résultats obtenus, plusieurs directions pourront être suivies, par exemple proposer des améliorations des algorithmes définis pendant la première année, résoudre des tâches plus complexes ou encore aborder des questions complémentaires. Le choix de la direction à suivre dépendra des difficultés et opportunités identifiées lors des premiers travaux avec pour objectif final la soutenance de la thèse et donc la réalisation d'une contribution scientifique significative et cohérente dans les 3 ans impartis.

3ème année

La troisième année sera consacrée essentiellement à la finalisation d'articles et à la rédaction du manuscrit de thèse. Elle pourra également être consacrée à des extensions des travaux réalisés précédemment dans l'objectif d'étayer au maximum le travail de thèse réalisé.

References

[1] Antoine Cully and Yiannis Demiris. Quality and diversity optimization: A unifying modular framework. *IEEE Transactions on Evolutionary Computation*, 22(2):245–259, 2017.

[2] Stephane Doncieux, Nicolas Bredeche, Léni Le Goff, Benoît Girard, Alexandre Coninx, Olivier Sigaud, Mehdi Khamassi, Natalia Díaz-Rodríguez, David Filliat, Timothy Hospedales, et al. Dream architecture: a developmental approach to open-ended learning in robotics. *arXiv preprint arXiv:2005.06223*, 2020.

[3] Stephane Doncieux, David Filliat, Natalia Díaz-Rodríguez, Timothy Hospedales, Richard Duro, Alexandre Coninx, Diederik M Roijers, Benoît Girard, Nicolas Perrin, and Olivier Sigaud. Open-ended learning: a conceptual framework based on representational redescription. *Frontiers in neurorobotics*, page 59, 2018.

[4] Seungsu Kim, Alexandre Coninx, and Stéphane Doncieux. From exploration to control: learning object manipulation skills through novelty search and local adaptation. *Robotics and Autonomous Systems*, 136:103710, 2021.

[5] Joel Lehman and Kenneth O Stanley. Abandoning objectives: Evolution through the search for novelty alone. *Evolutionary computation*, 19(2):189–223, 2011.

[6] Timothée Lesort, Natalia Díaz-Rodríguez, Jean-Franois Goudou, and David Filliat. State representation learning for control: An overview. *Neural Networks*, 108:379–392, 2018.

[7] Bryan Lim, Luca Grillotti, Lorenzo Bernasconi, and Antoine Cully. Dynamics-aware quality-diversity for efficient learning of skill repertoires. *arXiv preprint arXiv:2109.08522*, 2021.

[8] Carlos Maestre, Antoine Cully, Christophe Gonzales, and Stephane Doncieux. Bootstrapping interactions with objects from raw sensorimotor data: a novelty search based approach. In *2015 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, pages 7–12. IEEE, 2015.

[9] A. Merckling, N. Perrin-Gilbert, A. Coninx, and S. Doncieux. Exploratory state representation learning. *Frontiers in Robotics and AI*, 2022. "Unsupervised Representation Learning in Robotics" Research Topic.

[10] Giuseppe Paolo, Alexandre Coninx, Stéphane Doncieux, and Alban Laflaquière. Sparse reward exploration via novelty search and emitters. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 154–162, 2021.

[11] Giuseppe Paolo, Alban Laflaquiere, Alexandre Coninx, and Stephane Doncieux. Unsupervised learning and exploration of reachable outcome space. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2379–2385. IEEE, 2020.

[12] Justin K Pugh, Lisa B Soros, and Kenneth O Stanley. Quality diversity: A new frontier for evolutionary computation. *Frontiers in Robotics and AI*, 3:40, 2016.