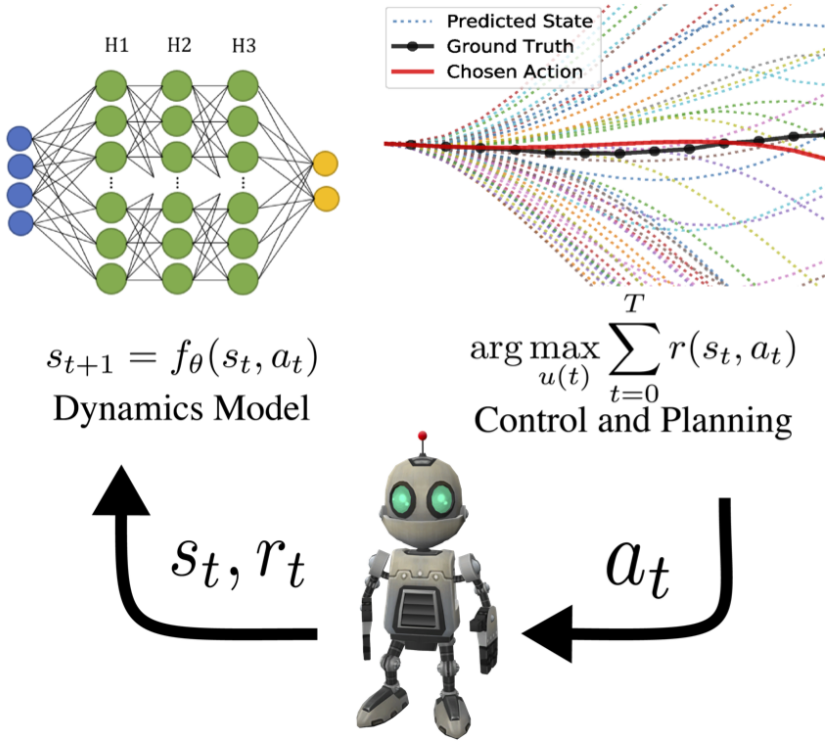
Apprentissage par renforcement guidé par la physique
Stage de Master 2 - 2022

1 Contexte

L'apprentissage profond par renforcement (*deep reinforcement learning*) a connu des succès impressionnants ces dernières années, par exemple au jeu de GO [13] ou pour contrôler des robots [12]. Les méthodes *model-free* [7, 12, 3, 6] permettent d'apprendre des politiques complexes en se basant uniquement sur des transitions entre états. Toutefois, leur généralité se fait au prix de la nécessité d'un très grand volume de données. Au contraire, les méthodes *model-based* [5, 10, 4, 17, 11, 8] exploitent un modèle de la dynamique de l'environnement pour un apprentissage plus efficace. Cependant, la qualité du modèle dynamique est souvent un frein à la performances de ces méthodes, qui reste asymptotiquement inférieure à celles des méthodes *model-free*. Pour pallier à cette déficience et limiter la dérive des trajectoires prédites, les méthodes actuelles ont tendance à considérer des trajectoires courtes et à replanifier fréquemment (model predictive control, MPC [10]).



2 Objectifs

Ce stage a pour objectif d'explorer l'utilisation de modèles physiques simplifiés en *model-based reinforcement learning (MBRL)*. Une piste d'étude intéressante est d'augmenter ces modèles simplifiés avec un modèle d'apprentissage permettant de compléter l'information manquante, par exemple un réseau de neurones (qui bénéficie des propriétés d'approximateurs universels de fonctions). Ce type d'approche a récemment été étudiée pour la prédiction de vidéos et de systèmes dynamiques [2, 9, 16], mais pas encore en RL.

L'objectif de ce stage sera d'adapter la méthode d'augmentation [16] au contexte de l'apprentissage par renforcement. Une réflexion sera menée concernant le choix du modèle simplifié de dynamique, qui pourrait être un modèle linéaire, localement linéaire ou un modèle physique plus dédié au problème, par exemple dont la dynamique est régie par une équation aux dérivées partielles ou ordinaire (ODE/PDE). Plusieurs bénéfices sont attendus d'un meilleur modèle dynamique dans le cadre de l'apprentissage par renforcement pour le contrôle. En particulier, ceci doit permettre un apprentissage plus rapide avec un nombre de simulations plus faible, ou encore la nécessité de replanifier moins souvent.

Différents cas d'usage pourront être explorés au cours de ce stage : nous commencerons par des problèmes "simples" contrôle en mécanique newtonienne (pendule simple, pendule double). En fonction de l'avancement du stage et des résultats, des tâches plus complexes seront abordées, comme des simulations de robotique (par exemple Reacher, Hopper avec la Deep Mind Control Suite [14]) ou bien des cas réels comme le contrôle de ballons stratosphériques [1] ou la nage de poissons [15].

3 Profil

Nous recherchons pour ce stage un·e candidat·e de niveau M2 ou dernière année d'école d'ingénieur avec une formation en mathématiques appliquées ou en informatique. Le ou la candidat·e idéal·e a une appétence pour la recherche scientifique et des bases théoriques en apprentissage automatique. Des notions en apprentissage par renforcement ou en systèmes dynamiques sont un plus pour ce sujet.

Une connaissance de la programmation avec Python est préférable, il est toutefois envisageable pour un·e candidat·e connaissant un autre langage de programmation de se former à Python au cours du stage. Une première expérience avec une bibliothèque d'apprentissage profond telle que TensorFlow ou PyTorch est la bienvenue. L'intérêt pour les librairies de simulation en robotique telles que Mujoco ou PyBullet serait un vrai plus.

4 Organisation

Cette offre de stage porte sur un stage d'une durée de 5 à 6 mois avec une date de début flexible au printemps 2022. Le stage se déroulera au centre de recherche et détudes en informatique et en communications (CEDRIC) du Conservatoire national des arts et métiers (Cnam) à Paris, 3e arrondissement.

Le CEDRIC est un laboratoire fondé en 1988 rassemblant plus de 80 enseignants-chercheurs regroupés dans 7 équipes thématiques. Ses activités couvrent divers champs de recherche allant de la fouille de données multimédia aux radiocommunications en passant par l'apprentissage statistique, les médias interactifs et loptimisation combinatoire.

Le stage sera co-encadré par Vincent Le Guen (EDF), Clément Rambour et Nicolas Thome de l'équipe Données complexes, apprentissage et représentations.

5 Candidater

Envoyer une candidature (CV + brève explication de votre motivation) par email à :
vincent.le-guen@edf.fr, clement.rambour@cnam.fr et nicolas.thome@cnam.fr.

Références

[1] M. G. Bellemare, S. Candido, P. S. Castro, J. Gong, M. C. Machado, S. Moitra, S. S. Ponda, and Z. Wang. Autonomous navigation of stratospheric balloons using reinforcement learning. *Nature*, 588(7836):77–82, Dec 2020.

[2] V. L. Guen and N. Thome. Disentangling physical dynamics from unknown factors for unsupervised video prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11474–11484, 2020.

[3] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine. Soft actor-critic : Off-policy maximum entropy deep reinforcement learning with a stochastic actor. *arXiv preprint arXiv:1801.01290*, 2018.

[4] D. Hafner, T. Lillicrap, I. Fischer, R. Villegas, D. Ha, H. Lee, and J. Davidson. Learning latent dynamics for planning from pixels. In *International Conference on Machine Learning*, pages 2555–2565, 2019.

[5] M. Hesse, J. Timmermann, E. Hüllermeier, and A. Trächtler. A reinforcement learning strategy for the swing-up of the double pendulum on a cart. *Procedia Manufacturing*, 24:15 – 20, 2018. 4th International Conference on System-Integrated Intelligence : Intelligent, Flexible and Connected Systems in Products and Production.

[6] A. X. Lee, A. Nagabandi, P. Abbeel, and S. Levine. Stochastic latent actor-critic : Deep reinforcement learning with a latent variable model. *arXiv preprint arXiv:1907.00953*, 2019.

[7] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra. Continuous control with deep reinforcement learning. *International Conference on Learning Representations*, 2016.

[8] M. Lutter, C. Ritter, and J. Peters. Deep lagrangian networks : Using physics as model prior for deep learning. In *International Conference on Learning Representations (ICLR 2019)*. OpenReview. net, 2019.

[9] V. Mehta, I. Char, W. Neiswanger, Y. Chung, A. O. Nelson, M. D. Boyer, E. Kolemen, and J. Schneider. Neural dynamical systems : Balancing structure and flexibility in physical prediction. *arXiv preprint arXiv:2006.12682*, 2020.

[10] A. Nagabandi, G. Kahn, R. S. Fearing, and S. Levine. Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7559–7566. IEEE, 2018.

[11] J. Schrittwieser, I. Antonoglou, T. Hubert, K. Simonyan, L. Sifre, S. Schmitt, A. Guez, E. Lockhart, D. Hassabis, T. Graepel, T. Lillicrap, and D. Silver. Mastering atari, go, chess and shogi by planning with a learned model, 2019.

[12] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz. Trust region policy optimization. In *International conference on machine learning*, pages 1889–1897. PMLR, 2015.

[13] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016.

[14] Y. Tassa, Y. Doron, A. Muldal, T. Erez, Y. Li, D. d. L. Casas, D. Budden, A. Abdolmaleki, J. Merel, A. Lefrancq, et al. Deepmind control suite. *arXiv preprint arXiv:1801.00690*, 2018.

[15] S. Verma, G. Novati, and P. Koumoutsakos. Efficient collective swimming by harnessing vortices through deep reinforcement learning. *Proceedings of the National Academy of Sciences*, 115(23):5849–5854, 2018.

[16] Y. Yin, V. L. Guen, J. Dona, E. de Bézenac, I. Ayed, N. Thome, and P. Gallinari. Augmenting physical models with deep networks for complex dynamics forecasting. *International Conference on Learning Representations*, 2021.

[17] M. Zhang, S. Vikram, L. Smith, P. Abbeel, M. Johnson, and S. Levine. Solar : Deep structured representations for model-based reinforcement learning. In *International Conference on Machine Learning*, pages 7444–7453, 2019.