Rapport prévisionnel

Développez une preuve de concept

Alan Blanchet

Tuteur Neovision : Arthur DERATHE
Mentor OpenClassrooms : Chemsse EDDINE NABTI

1 Introduction

Ce rapport prévisionnel présente les recherches que j'ai mené pour le choix de mon sujet *Proof of Concept* selon les critères d'OpenClassrooms.

Ayant peu de connaissances en Reinforcement Learning, ce projet est un challenge pour moi et je dois donc tout d'abord me former pour mieux maitriser celui-ci.

J'ai reccueilli toutes les informations que j'ai pu trouver sur le sujet dans un répertoire GitHub [1]

2 Recherches

Je me suis d'abord renseigné sur les bases du Reinforcement Learning. En commencant une simple matrice d'état à action Q, puis sur les différents algorithmes existants plus récents.

J'ai réimplémenté l'algorithme REINFORCE[3] classique qui fait déjà usage du Deep Learning. Puis j'ai étudié les algorithmes VPG[3]/TRPO[2]/PPO qui sont des améliorations de celui-ci.

Je me suis également renseigné à partir de nombreux repositories Github sur les implémentations plusieurs algorithmes. Le site d'OpenAI est également une source d'information importante car ce sont eux-même qui ont développé une solution pour abstraire la notion d'environnement de simulation afin d'effectuer des tests sur différents jeux.

3 Algorithme retenu

Après avoir compris une multitude de notions différentes. Je me suis mis à la recherche d'un algorithme récent qui pourrait satisfaire les contraites du projet.

J'ai donc choisi l'algorithme NGU[1] (2020) qui utilise un mechsanisme de curiosité pour améliorer l'apprentissage de l'agent.

1. https://github.com/AlanBlanchet/AI-4-Alan

Références

[1] Adrià Puigdomènech Badia, Pablo Sprechmann, Alex Vitvitskyi, Daniel Guo, Bilal Piot, Steven Kapturowski, Olivier Tieleman, Martín Arjovsky, Alexander Pritzel, Andew Bolt et Charles Blundell. *NEVER GIVE UP: LEARNING DIRECTED EXPLORATION STRATEGIES*. https://openreview.net/pdf?id=Sye57xStvB. 2020.

[2] John Schulman, Sergey Levine, Philipp Moritz, Michael I. Jordan et Pieter Abbeel. *Trust Region Policy Optimization*. https://arxiv.org/pdf/1502.05477.pdf. 2017.

[3] Richard S. Sutton, David McAllester, Satinder Singh et Yishay Mansour. *Policy Gradient Methods for Reinforcement Learning with Function Approximation*. https://proceedings.neurips.cc/paper_files/paper/1999/file/464d828b85b0bed98e80adePaper.pdf. 1999.