Ecole Normale Supérieure de Paris-Saclay

Rapport TER

TER - Voiture autonomes avec apprentissage par renforcement et lidar

---

14 mai 2024

Miquel Hugo

Plus Basile

Table des matières

1 Introduction

1.1 Contexte

Les voitures autonomes sont un sujet de recherche très actif depuis quelques années. En effet, elles pourraient révolutionner le monde des transports en permettant de réduire les accidents de la route, de diminuer la consommation d'énergie et de réduire les embouteillages. Cependant, il reste encore de nombreux défis à relever pour que les voitures autonomes soient utilisées à grande échelle. En particulier, il est nécessaire de développer des algorithmes d'apprentissage par renforcement qui permettent à une voiture autonome d'apprendre à conduire de manière autonome.

1.2 Objectif et travail réalisé

L'objectif de ce TER est de développer un algorithme d'apprentissage par renforcement qui permet à une voiture RC au format $1/10^{\text{ème}}$ de conduire de manière autonome sur un circuit. Dans un premier temps, nous avons utilisé Webots pour la simulation, gym et stable baselines pour l'apprentissage par renforcement. Dans un second temps, nous avons transféré le réseau de neurones du simulateur à la voiture réelle. La voiture est équipée d'un lidar qui permet de mesurer la distance entre la voiture et les murs du circuit.

Note : Presenter Webots, la voiture réelle, la voiture sur simulateur, le lidar, le circuit etc...

1.3 L'apprentissage par renforcement

L'apprentissage par renforcement est une méthode d'apprentissage automatique qui permet à un agent d'apprendre à prendre des décisions en interagissant avec un environnement.

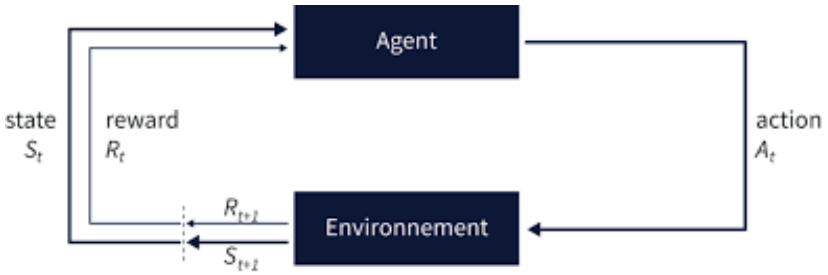

FIGURE 1 – Schéma de l'apprentissage par renforcement

L'agent prend des actions dans l'environnement et reçoit une récompense en fonction de l'action qu'il a prise. L'objectif de l'agent est de maximiser la somme des récompenses qu'il reçoit au cours des itérations.

Lors de chaque étape, l'agent reçoit une observation de l'environnement dans lequel il évolue. Sur la base de cette observation, l'agent prend une décision parmi un ensemble d'actions possible appelé espace des actions. Cet espace peut dépendre de l'état dans lequel se trouve l'agent.

Un exemple simple est celui d'un jeu d'échecs dans lequel l'observation correspond à la position de chacune des pièces sur l'échiquier et l'espace des actions est l'ensemble des déplacements possibles des pièces. Naturellement, on souhaite que l'agent réalise la meilleure action possible suivant l'observation reçue. Pour atteindre ce but, l'agent

applique une politique d'action (notée π) qu'il utilise pour sa prise de décision. À chaque récompense obtenue, cette politique est mise à jour. On espère ainsi atteindre une politique optimale.

Pour entraîner un agent, plusieurs types de méthodes peuvent être utilisées. Ces méthodes estiment la somme des récompenses futures que l'agent devrait obtenir. Ces récompenses sont pondérées pour favoriser les récompenses à court terme. La politique obtenue est souvent modélisée par un réseau de neurones, dont l'actualisation modifie les poids du réseau.

Les méthodes d'apprentissage par renforcement peuvent être classées en trois catégories principales :

- **Méthodes basées sur la valeur (value-based)** : Ces méthodes se concentrent sur l'estimation de la récompense cumulative optimale que l'agent peut obtenir. Elles cherchent à obtenir une récompense cumulative maximale.
- **Méthodes basées sur la politique (policy-based)** : Ces méthodes se concentrent sur l'optimisation de la politique de l'agent. Les valeurs de récompense peuvent ne pas être calculées directement.
- **Méthodes acteur-critique (actor-critic)** : Ces méthodes utilisent deux réseaux de neurones. Le premier réseau choisit l'action à effectuer, tandis que le second réseau évalue cette action en la comparant à l'action prévue.

2 Voiture

3 Simulation

3.1 Le simulateur Webots

Le logiciel utilisé pour la simulation est Webots R2023b. Webots est un logiciel de simulation de robotique développé par Cyberbotics. Il permet de simuler des robots dans un environnement 3D que l'on peut personnaliser. Dans notre cas, nous avons utilisé Webots pour simuler une voiture RC sur un circuit. Nous avons utilisé le langage de programmation Python pour contrôler la voiture dans le simulateur.

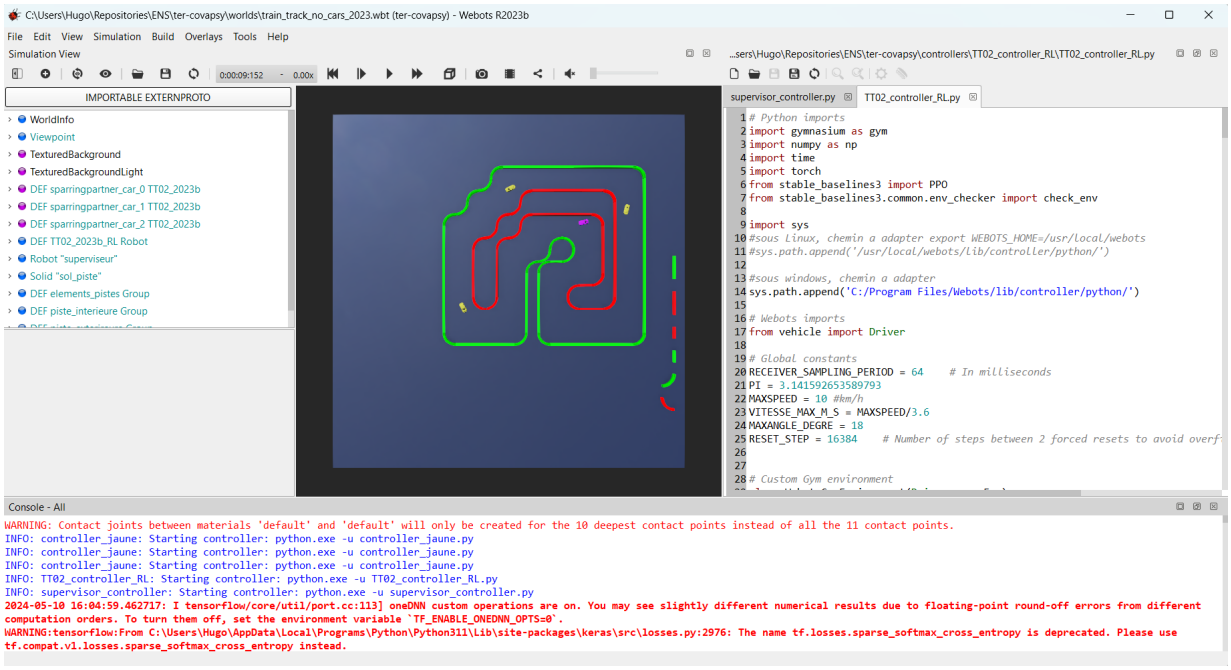

FIGURE 2 – Capture d'écran de Webots

Note : Présenter le circuit, la voiture, le lidar, le code Python etc...

3.2 Le circuit

Le circuit dans l'environment de Webots a pour but de simuler un circuit réel sur lequel la voiture autonome doit apprendre à conduire. Les murs du circuit sont composés de blocs de couleur différentes pour les bordures extérieur et intérieur du circuit. Ces murs ont une auteur d'une dizaine de centimètres qui permettent au lidar de mesurer la distance entre la voiture et les murs.

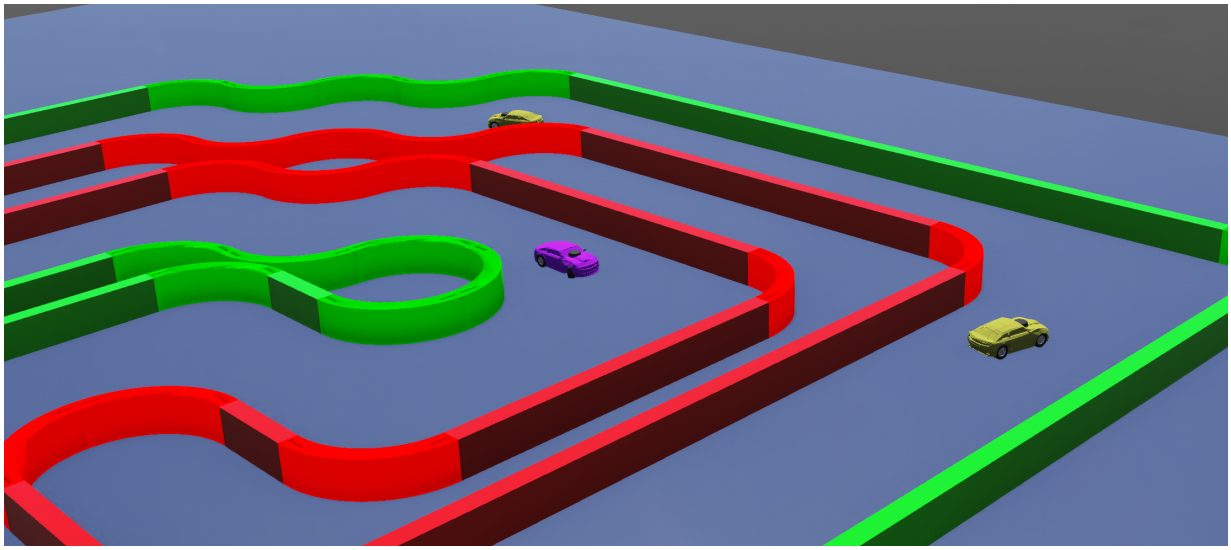

FIGURE 3 – Capture d'écran du circuit

3.3 La voiture

La voiture utilisée dans le simulateur est une voiture RC au format $1/10^{\text{ème}}$. Elle a pour objectif de reproduire le plus fidèlement possible une voiture réelle. La voiture est équipée d'un lidar qui permet de mesurer la distance entre la voiture et les murs du circuit.

FIGURE 4 – Capture d'écran de la voiture

3.4 Le lidar

4 Simulation to real world

L'objectif du SimToReal est de transférer un réseau de neurones entraîné sur un simulateur à une voiture réelle. Une fois le réseau de neurones entraîné, il est transféré à la voiture réelle pour qu'elle puisse conduire de manière autonome sur un circuit réel. Un des principaux défis du SimToReal est de reproduire le plus fidèlement possible les conditions du monde réel dans le simulateur.

La phase d'entraînement sur simulateur a plusieurs avantages. Tout d'abord, elle permet de réduire le temps et le coût de l'entraînement. En effet, il est possible de simuler des milliers d'épisodes en quelques heures alors qu'il faudrait plusieurs jours pour réaliser le même nombre d'épisodes sur une voiture réelle. De plus, la simulation permet de tester des scénarios dangereux pour la voiture réelle sans risquer de l'endommager. Sur la voiture réelle, il faut aussi la replacer à la main après chaque crash dans un mur. Il est donc plus efficace de réaliser l'entraînement sur simulateur.

5 Introduction du bruit pour améliorer la simulation

Pour que la simulation soit la plus proche possible de la réalité, il est crucial d'introduire des éléments de bruit. Ces bruits peuvent affecter les mesures et les actions de la voiture, et ainsi permettre au réseau de neurones de mieux généraliser lors du passage au monde réel.

5.1 Bruit sur les mesures

Dans un environnement réel, les capteurs ne fournissent pas des mesures parfaites. Les mesures du Lidar peuvent toujours être affectées par de petites interférences ou des variations mineures. Pour simuler ces conditions, on peut ajouter un léger bruit gaussien aux mesures du Lidar. Cela permet au réseau de neurones de s'adapter à des données moins parfaites, similaires à celles qu'il rencontrera dans le monde réel.

5.2 Bruit sur les actions

Les actions de la voiture, telles que l'angle de direction ou la vitesse, peuvent également être sujettes à des variations imprévues. Par exemple, un servomoteur peut ne pas toujours répondre de manière identique à une même commande en raison de l'usure ou des variations de tension. Pour prendre en compte ces imperfections, on peut ajouter du bruit aux actions commandées par le réseau de neurones. Cela aide à rendre l'agent plus robuste face aux variations qu'il pourrait rencontrer sur une voiture réelle.

6 Application à la voiture autonome

Pour appliquer l'apprentissage par renforcement à une voiture autonome, il est essentiel de définir correctement l'espace d'observation et l'espace d'action en tenant compte des contraintes du monde réel.

6.1 Espace d'observation

L'espace d'observation doit inclure toutes les informations pertinentes que la voiture peut obtenir de son environnement. Dans notre cas, nous utilisons le Lidar pour mesurer les distances aux obstacles. Si un système de contrôle de la vitesse est en place, la vitesse actuelle de la voiture pourrait également faire partie de l'espace d'observation.

6.2 Espace d'action

L'espace d'action est limité aux commandes que l'agent peut envoyer à la voiture. Cela inclut l'incrémentation de l'angle de direction via le servomoteur et l'incrémentation de la vitesse via le moteur.

7 Conclusion