

# Ghazi Bel Mufti

---

Titre :

Le contrôle d'un système dynamique vu sous le prisme du reinforcement learning

Résumé :

La théorie du contrôle analyse les propriétés des systèmes commandés, c'est-à-dire des systèmes dynamiques sur lesquels on peut agir au moyen d'une commande (ou contrôle). Le but est alors d'amener le système d'un état initial donné à un certain état final, en respectant éventuellement certains critères. L'objectif peut être de stabiliser le système pour le rendre insensible à certaines perturbations (stabilisation), ou encore de déterminer des solutions optimales pour un certain critère d'optimisation (contrôle optimal).

Nous nous intéressons aux problèmes de contrôle optimal de systèmes dynamiques, vus sous le prisme de l'apprentissage par renforcement. L'apprentissage par renforcement dérive de formalisations théoriques de méthodes de contrôle optimal, visant à mettre au point un contrôleur permettant de minimiser au cours du temps une mesure donnée du comportement d'un système dynamique. La version discrète et stochastique de ce problème est appelée un processus de décision markovien et fut introduite par Bellman en 1957. Parmi les premiers algorithmes d'apprentissage par renforcement, on compte le Temporal Difference learning (TD-learning), proposé par Richard Sutton (1988), et le Q-learning mis au point essentiellement par Chris Watkins (1992).

En intelligence artificielle, et plus précisément en apprentissage automatique, l'apprentissage par renforcement consiste, pour un agent autonome (robot, etc.), à apprendre les actions à prendre, à partir d'expériences, de façon à optimiser une récompense quantitative au cours du temps. L'agent est plongé au sein d'un environnement, et prend ses décisions en fonction de son état courant. En retour, l'environnement procure à l'agent une récompense, qui peut être positive ou négative. L'agent cherche, au travers d'expériences itérées, un comportement décisionnel (appelé stratégie ou politique, et qui est une fonction associant à l'état courant l'action à exécuter) optimal, en ce sens qu'il maximise la somme des récompenses au cours du temps.

Dans de nombreux problèmes pratiques de prise de décision, les états du processus de décision Markovien sont de grande dimension (par exemple, les images d'une caméra ou le flux de capteur brut d'un robot) et ne peuvent pas être résolus par les algorithmes d'apprentissage par renforcement traditionnels. Les algorithmes d'apprentissage par renforcement profond intègrent un apprentissage profond pour résoudre de tels problèmes, représentant souvent la politique ou d'autres fonctions apprises telle que la fonction Q en tant que réseau de neurones, et développant des algorithmes spécialisés qui fonctionnent bien dans ce cadre. L'apprentissage par renforcement profond (Deep Reinforcement Learning DRL) est un sous-domaine de l'apprentissage automatique qui combine l'apprentissage par renforcement (RL) et l'apprentissage profond (Deep Learning). L'apprentissage par renforcement profond intègre l'apprentissage profond dans la solution, permettant aux agents de prendre des décisions à partir de données d'entrée non structurées sans

ingénierie manuelle de l'espace d'état. Les algorithmes d'apprentissage par renforcement profond sont capables de prendre en compte de très grandes entrées (par exemple, chaque pixel affiché à l'écran dans un jeu vidéo) et de décider des actions à effectuer pour optimiser un objectif (par exemple, maximiser le score du jeu).

Après une brève présentation des concepts de RL et DRL, nous présentons une nouvelle approche qui combine ces concepts avec les outils de contrôle optimal dans le cas linéaire ainsi qu'une application de cette approche à la résolution du problème du pendule inversé.

Références :

1. Changsheng Hua, Reinforcement Learning Aided Performance Optimization of Feedback Control Systems, 2020.
2. Richard Sutton and Andrew Barto, Reinforcement learning : an introduction. MIT Press, 2018.
3. Matthieu Geist, Contrôle optimal et apprentissage automatique, applications aux interactions homme-machine, 2016.
4. Yi Ma and Shankar Sastry, Reinforcement Learning & Optimal Control Overview, University of California, Berkeley, 2021.
5. Swagat Kumar, Balancing a CartPole System with Reinforcement Learning -- A Tutorial, arXiv preprint, arXiv:2006.04938, 2020.