

3SQ2040 - Apprentissage par renforcement

Responsables : **Richard COMBES**

Langues d'enseignement : **ANGLAIS**

Campus o  le cours est propos  : **CAMPUS DE PARIS - SACLAY**

Nombre d'heures d' tudes  l ves (HEE) : **40**

Nombre d'heures pr sentielles d'enseignement (HPE) : **21**

Ann e acad mique : **2024-2025**

Niveau avanc  : **non**

Pr sentation, objectifs g n raux du cours :

L'apprentissage par renforcement est un domaine de l'intelligence artificielle qui traite les m thodes qui permettent pour apprendre par essais et erreurs. L'apprentissage par essais et erreur est non seulement la strat gie utilis e par tous les  tres vivants mais  galement celle qui permet de r soudre efficacement plusieurs probl mes difficiles et importants:

- comment apprendre   un ordinateur   jouer aux  checs ou au jeu de go et battre les meilleurs joueurs humains ?

- comment apprendre   des robots   se mouvoir dans un environnement impr vu,  viter les obstacles et atteindre des cibles ?

- comment recommander automatiquement des contenus (livres, films, vid os etc.)   des millions de clients, afin que chacun obtienne des recommandations qui l'int resse ?

Le but de ce cours est de donner une introduction   ce domaine   la fois fascinant et en pleine expansion: des mod les math matiques, en passant par les algorithmes, leur performance th orique et num rique, ainsi que leur impl mentation pratique.

P riode(s) du cours (n  de s quence ou hors s quence) :

SM11

Pr requis :

Cours de premi re ann e de probabilit s et statistiques. Niveau basique de programmation en Python.

Plan d taill  du cours (contenu) :

Cours 1:

Partie 1: Mod le g n ral de l'apprentissage par renforcement:  tats, actions, r compenses, exemples d'applications.

Partie 2: Probl mes de Bandits I: d finitions, r compenses stochastiques vs adversariales, algorithme EXP3.

Cours 2:

Partie 1: Probl mes de Bandits II: borne de Lai-Robbins, algorithme Upper Confidence Bound, algorithme Thompson sampling, probl mes structur s.

Partie 2: Probl mes de Bandits III: impl mentation en Python et performance en pratique.

Cours 3:

Partie 1: Processus de Décision Markoviens I: définitions, équation de Bellman, calcul des politiques optimales.

Partie 2: Processus de Décision Markoviens II: résoudre les Processus de Décision Markoviens en ligne, algorithme Q-learning et plus.

Cours 4:

Partie 1: Processus de Décision Markoviens III: fléau de la dimension , approximation de fonctions valeur, gradient de politique .

Partie 2: Processus de Décision Markoviens IV: implémentation en Python et performance en pratique.

Déroulement, organisation du cours :

Cours magistraux + programmation en Python.

Organisation de l'évaluation :

Compte rendu à rendre à la fin du cours.

Moyens :

Cours magistraux + programmation en Python.

Acquis d'apprentissage visés dans le cours :

A l'issue de ce cours, les étudiants maîtriseront les algorithmes principaux de l'apprentissage par renforcement. Ils comprendront leur garanties théoriques de performance, ainsi que comment les implémenter et leur performance pratique.

Description des compétences acquises à l'issue du cours :

With this course, students will have a firm grasp of the main algorithms used in reinforcement learning. They will understand their theoretical performance guarantees, as well as how to implement, and their actual performance.

Bibliographie :

Notes de cours