UNIVERSITÉ DE ROUEN

M2 SD, Année 2020-2021
Jeudi 3 Décembre 2020

Méthodes de Monte Carlo, Différence Temporelle

1 Présentation du problème

Nous allons appliquer les algorithmes d'*évaluation de la value-function* à partir d'épisodes générés dans leur intégralité ou en ligne (*Temporal Difference*) dans le cadre de l'exploration d'un labyrinthe. Comme dans le cadre du TP précédent, cet espace sera représenté sous forme de grille et le point de départ d'une expérience sera une des cases correspondant à l'intérieur du labyrinthe et les actions possibles seront données en fonction des directions correspondant à une case vide adjacente. Une seule sortie existe et l'arrivée sur cette sortie correspond à une récompense de 1, tout autre mouvement étant accompagné d'une récompense nulle.

Tous ces éléments sont fournis : le fichier [1] `Maze_generating_interface.py` permet de dessiner un labyrinthe et le fichier `ForTP3.py` permet de générer des épisodes complet (se terminant à la sortie) ou d'une longueur fixe, dans le cadre de la politique aléatoire uniforme.

2 Objectifs

2.1 Monte-Carlo methods

Votre objectif est d'implémenter les algorithmes *First-Visit MC prediction*,*Monte Carlo Exploring Starts*, *On-policy first-visit MC control* dans le cadre d'un Labyrinthe de taille 15*15.
Vous pouvez tester les choses sur des instances plus petites, les épisodes générés par la politique aléatoire peuvent être particulièrement long.

2.2 Temporal Difference methods

Votre objectif est d'implémenter les algorithmes *TD[0]*,*SARSA* dans le cadre du même Labyrinthe de taille 15*15.

3 Rendus

— Le code des méthodes dans un fichier .py
— Un compte rendu en .pdf ou un notebook.

4 Bonus

Mise en place du $n - step$

1. issu d'un projet mené par O. Nanushi