

TP noté : Apprentissage par Renforcement - Q-Learning (GridWorld)

Objectifs

- Visualiser la politique apprise
- Analyser l'influence des hyperparamètres
- Tester la capacité de généralisation d'un tel modèle

Contexte

Pour ce TP, nous allons considérer à nouveau un agent qui évolue dans une grille 5x5 et doit atteindre un objectif en évitant des pièges.

L'objectif est matérialisé par une case spécifique.

De même, les pièges sont des cases particulières dans son environnement, c'est à dire la grille 5x5.

Dans ce contexte, l'état perçu par l'agent est sa position dans la grille.

Il possède 4 actions : aller vers le haut, vers le bas, vers la gauche ou vers la droite.

Vous avez complété le programme correspondant lors de la dernière séance ou bien vous pourrez repartir du programme complet mis en ligne.

Travail demandé

Pour chaque point demandé ci-dessous, vous veillerez à implémenter le cas échéant les éléments pas à pas en réalisant des tests au fur et à mesure.

Vous rendrez en fin de séance vos **programmes** ainsi qu'un **compte rendu d'expériences** avec une **analyse critique** des résultats.

1. Complétez votre programme pour **visualiser** après entraînement la **politique apprise** par l'agent c'est-à-dire l'action à mener suivant chaque état.
2. Analysez les résultats et l'influence des **hyperparamètres** sur la vitesse de convergence et la politique optimale :

- a. Epsilon, alpha, gamma
- b. Récompenses
- c. Nombre de pièges
- d. Taille de la grille

3. Analysez la **capacité de généralisation** du modèle face aux éléments suivants :

- a. Changer la de position du but,
- b. Rajouter un piège, plusieurs pièges
- c. Changer la position des pièges
- d. Changer le nombre de pièges
- e. Agrandir la taille de l'environnement

Pour cela, vous estimerez les performances sans ré-entraînement ainsi que le nombre d'épisodes d'entraînements nécessaires pour atteindre l'objectif dans chacun de ces nouveaux contextes

Complétez vos programmes en conséquence.