

TP noté : Apprentissage par Renforcement - Q-Learning (“chien et moutons”)

Objectifs

- Analyser l'influence des hyperparamètres de la simulation et de l'apprentissage
- Tester la capacité de généralisation d'un tel modèle
- Proposer des descripteurs différents

Contexte

Pour ce TP, nous considérons un monde dans lequel un chien, l'agent qui va apprendre, essaie de « pousser » les moutons présents dans le champ dans un enclos.

La **grille** ici est de taille 10 x 10 et le nombre de moutons est un paramètre du programme.

Un **épisode** est terminé quand tous les moutons sont dans l'enclos.

L'espace des états étant trop grand pour une approche d'apprentissage par renforcement tabulaire (ici le Q-Learning), nous avons choisi que la perception du chien soit définie par 4 **descripteurs** :

1. La *distance* du *centre de gravité des moutons* dans l'herbe à la porte de l'enclos
2. La *distance* du *chien au centre de gravité des moutons dans l'herbe*
3. Un descripteur prenant en compte *l'angle* entre la direction de la porte et la direction du centre de gravité des moutons

Ces descripteurs sont une sorte de compression de la situation du monde, et donc conduisent à des états partiellement observables puisque plusieurs états du monde peuvent être perçus comme identiques par le chien.

Il peut réaliser 4 **actions** :

1. Se *déplacer* en direction du *centre de gravité du troupeau dans l'herbe*
2. Se *déplacer* latéralement vers la *gauche* par rapport à cette direction
3. Se *déplacer* latéralement vers la *droite* par rapport à cette direction
4. Il reste sur place et *aboie*.

Vous disposez d'un programme permettant de réaliser des expériences en modifiant les paramètres de la simulation et de l'apprentissage.

Travail demandé

Pour chaque question ci-dessous, réalisez un protocole expérimental rigoureux et reproductible, sans imposer cependant de reproduire les valeurs tirées aléatoirement.

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

1. En partant d'une simulation **avec 2 moutons**, comment évolue la performance en termes de `reward` ? Combien d'épisodes semblent nécessaires pour atteindre une performance qui paraisse asymptotique ?
2. Faites **varier les paramètres de l'apprentissage** : alpha, gamma, epsilon de début et de fin, et reportez les performances obtenues. Conservez les valeurs qui semblent conduire aux meilleures performances.
3. Est-ce que l'évolution des performances est différente **si vous conservez la table des q-valeurs** entre deux apprentissages ?
4. Fixez maintenant le **nombre de moutons à 5**. Est-ce que vous observez des différences dans l'entraînement ? Est-ce que l'utilisation de la table des q-valeurs apprise pour 2 moutons facilite l'apprentissage pour 5 moutons ?
5. **Comparez la table des q-valeurs** après un apprentissage avec 2 moutons et avec 5 moutons. Identifiez-vous des points communs entre les deux tables ? Si oui, pouvez-vous comprendre à quoi correspondent les cases de ces tables qui sont similaires.
6. Même chose quand on réalise deux apprentissages avec des **positions initiales du chien** différentes.
7. Idem en modifiant la **position de l'enclos**.
8. On va maintenant s'attacher à **la perception du chien**. Modifiez les **paramètres** des descripteurs en essayant d'améliorer les performances.
9. Vous pouvez **proposer des descripteurs** différents qui vous sembleraient intéressants. Expérimentez et comparez les performances obtenues en les utilisant.