Faculté des Nouvelles Technologies de l'Information et de la Communication
Département : Informatique Fondamentale et ses Applications
Année Universitaire : 2023/2024
Module : DAIIA (Master 1 SDIA)
Enoncé TD N° 4

❖ Exercice 1:

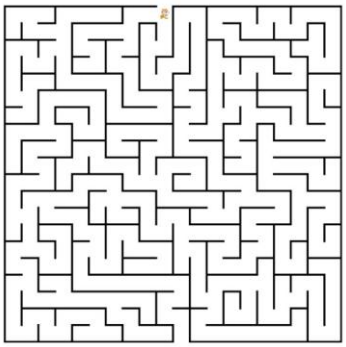
Supposons un problème de classification d'images où il y'a un ensemble de données d'images de fruits. Chaque image est associée à une étiquette indiquant le type de fruit qu'elle représente. Vous avez trois approches d'apprentissage que vous pourriez utiliser : l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage par renforcement.

1. Discutez de la manière dont vous pourriez aborder ce problème avec chacune de ces approches.
2. Discutez les avantages et les inconvénients de chacune.

❖ Exercice 2:

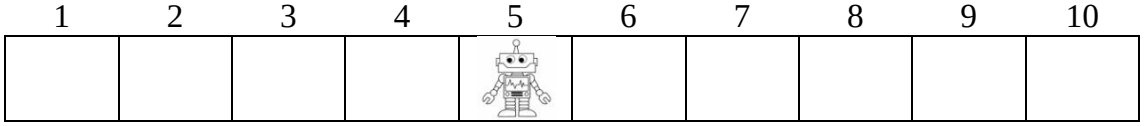
Considérez un robot mobile placé dans un labyrinthe rectangulaire. L'objectif du robot est de trouver le chemin le plus court pour atteindre une sortie spécifique du labyrinthe. Le robot peut prendre quatre actions : avancer d'une case, reculer d'une case, tourner à gauche et tourner à droite. Le labyrinthe est représenté comme une grille discrète où certaines cases sont des murs infranchissables, et le robot reçoit une récompense positive lorsqu'il atteint la sortie et une récompense négative s'il percute un mur.

- Formulez ce problème en utilisant le formalisme de l'apprentissage par renforcement.



❖ Exercice 3:

Considérons un agent qui doit apprendre à jouer à un jeu de plateau simple. Le plateau est composé de cases numérotées de 1 à 10, et l'agent commence au milieu (case 5). L'objectif est d'atteindre la case 10 pour maximiser la récompense.



1. Définissez le problème comme un processus décisionnel de Markov (MDP). Identifiez les éléments clés tels que l'espace d'états, l'espace d'actions, la fonction de transition, la fonction de récompense, et l'horizon temporel.
2. Proposez une politique initiale pour l'agent. Cette politique peut être déterministe ou stochastique.
3. Expliquez le concept d'exploration et d'exploitation dans le contexte de cet exemple. Comment l'agent peut-il équilibrer ces deux aspects pour améliorer son apprentissage?

❖ **Solution exercice 1:**

	Apprentissage supervisé	**Apprentissage non supervisé**	**Apprentissage par renforcement**
Approche	Utilisez un ensemble de données étiqueté pour entraîner un modèle qui peut prédire le type de fruit à partir d'une nouvelle image.	Utilisez des techniques d'apprentissage non supervisé telles que le regroupement (clustering) pour découvrir des motifs ou des similarités entre les images de fruits sans utiliser d'étiquettes.	L'agent (le modèle) interagit avec l'environnement (les images de fruits) et apprend à prendre des décisions (choisir le type de fruit) pour maximiser une récompense (classification correcte).
Avantages	Précision élevée, sortie clairement définie, bon pour la classification lorsque les étiquettes sont disponibles.	Peut identifier des structures sous-jacentes dans les données sans nécessiter d'étiquettes, utile pour l'exploration des données.	Peut être utilisé dans des environnements dynamiques, apprend de l'expérience, peut s'adapter à des situations nouvelles.
Inconvénients	Besoin d'un ensemble de données étiqueté volumineux, coût de l'annotation en temps et en ressources, ne fonctionne pas bien si les étiquettes sont bruitées ou inexistantes.	Les résultats peuvent être moins interprétables, dépend largement des algorithmes de regroupement choisis, peut ne pas être aussi précis que l'apprentissage supervisé.	Requiert une exploration judicieuse, peut être difficile à mettre en œuvre, nécessite un retour d'information (récompense) approprié.

❖ **Solution exercice 2:**

1. Espace d'État (State Space) : L'état de l'environnement est défini par la position actuelle du robot dans la grille du labyrinthe.
Ainsi, l'état **s est représenté par les coordonnées (x, y) du robot.**

2. Espace d'Actions (Action Space) : L'espace d'actions A comprend les quatre actions possibles pour le robot :
Avancer d'une case, reculer d'une case, tourner à gauche, et tourner à droite.
Donc :
A = {"Avancer", "Reculer" , "Tourner à gauche", "Tourner à droite"}

3. Fonction de transition (Transition Function) : La fonction de transition P détermine la probabilité de passer d'un état à un autre en prenant une action spécifique.
P(s' | s, a)

Par exemple, si le robot est à la position (x, y) et choisit l'action "Avancer", la fonction de transition déterminera la nouvelle position (x', y'), en prenant en compte les murs et les limites du labyrinthe.

4. Récompenses (Reward Function) : La fonction de récompense R attribue une récompense au robot en fonction de l'état actuel et de l'action prise.
R(s, a, s')

Une récompense positive est donnée lorsque le robot atteint la sortie (+5), et une récompense négative (-1) est donnée s'il percute un mur, et une récompense neutre ou légèrement négative (par exemple, 0 ou -0.01) pour chaque mouvement, pour encourager la découverte du chemin le plus court.

5. Politique (Policy) : La politique π est la stratégie que le robot suit pour choisir ses actions dans chaque état.

π (a | s)

✓ En apprentissage par renforcement, une politique (policy en anglais) représente la stratégie que l'agent utilise pour prendre des décisions dans un environnement donné.

✓ Formellement, une politique est une fonction qui attribue une action à chaque état possible de l'agent dans l'environnement. La politique est souvent notée : π.

N.B : le problème est formulé comme un processus de décision de Markov (MDP) défini par un ensemble d'états, d'actions, d'une fonction de transition, d'une fonction de récompense, d'une politique. L'objectif du robot est de trouver une politique optimale pour atteindre la sortie du labyrinthe tout en maximisant la somme des récompenses sur le temps T.

❖ **Solution exercice 3:**

1. MDP :

- Espace d'états (S) : {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
- Espace d'actions (A) : {Gauche, Droite}
- Fonction de transition (P) : Probabilité de passer à un nouvel état après avoir pris une action.
- Fonction de récompense (R) : Récompense associée à chaque transition état-action.
- Horizon temporel : Indique la durée maximale de l'épisode.

2. Politique Initiale :

- Une politique possible pourrait être de toujours aller à droite, c'est-à-dire choisir l'action "Droite" à chaque état.

3. Exploration et Exploitation :

- L'agent doit équilibrer l'exploration (essayer de nouvelles actions) et l'exploitation (choisir les actions les plus prometteuses).

Par exemple : La Stratégie ∈-Greedy

L'agent pourrait choisir une action exploratoire avec une probabilité epsilon « exploration » et choisir l'action la plus prometteuse avec une probabilité 1 – epsilon « exploitation ».

4. Fonction de Récompense :

- Une fonction de récompense pourrait attribuer une récompense positive lorsque l'agent atteint la case 10 et une récompense nulle ou négative sinon. Cela motive l'agent à maximiser les récompenses cumulées en atteignant la case