

❖ **Exercice – Pac-Man**

Pac-Man doit naviguer à travers un labyrinthe pour manger toutes les pac-gommes tout en évitant les fantômes. Chaque action prise par Pac-Man (déplacer vers le haut, le bas, la gauche ou la droite) a des conséquences sur son environnement, et il reçoit des récompenses en fonction de ses actions et de l'état du jeu. Les récompenses et les probabilités de transition sont les suivantes :

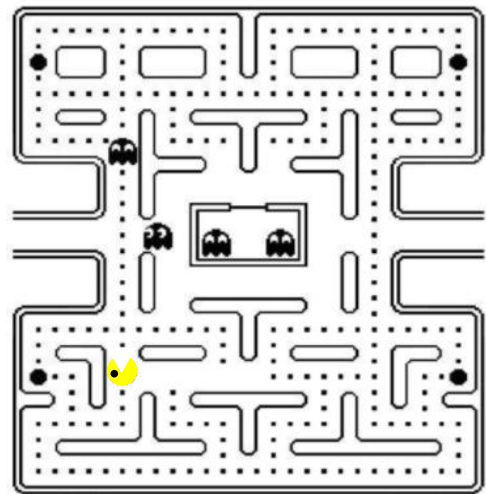
• **Récompenses :**

- Manger une pac-gomme : +10
- Être capturé par le fantôme : -50
- Atteindre la sortie du labyrinthe (gagner) : +100

• **Probabilités de transition :**

Pac-Man prend une action (déplacer vers le haut, le bas, la gauche ou la droite) :

- 80% de chances de se déplacer dans la direction choisie.
- 20% de chances de rester immobile.



Questions

1. Qu'est-ce que la fonction de valeur
2. Quelle est la différence entre la valeur d'un état et la valeur d'une action ?
3. Dans un état où Pac-Man est à côté d'une pac-gomme, si la fonction de valeur d'état est de $V(s) = 25$, quelle serait la valeur d'état mise à jour après avoir mangé la pac-gomme ?
4. Si Pac-Man peut se déplacer dans un état voisin et la récompense attendue est de 10, quelle serait la fonction de valeur d'état mise à jour après avoir pris une certaine action ?
5. Supposons que dans un état où Pac-Man peut se déplacer vers le haut, la récompense immédiate de prendre cette action est de +5. Si la valeur d'état après avoir pris cette action est de $V(s') = 30$, quelle serait la valeur d'action mise à jour pour cette action ?
6. Si les valeurs d'action pour un état donné sont $Q(s, \text{haut}) = 12$, $Q(s, \text{bas}) = 15$, $Q(s, \text{gauche}) = 18$ et $Q(s, \text{droite}) = 20$, comment pouvez-vous calculer la fonction de valeur d'état pour cet état ?
7. Comment la table Q est-elle structurée dans le contexte de l'apprentissage par renforcement de Pac-Man, et quelles informations sont stockées dans cette table ?

❖ Solution exercice Pac-Man :

1. Qu'est-ce que la fonction de valeur

C'est une estimation de la récompense attendue à long terme à partir d'un certain état ou d'une certaine action.

2. Quelle est la différence entre la valeur d'un état et la valeur d'une action ?

La valeur d'état offre une perspective globale de l'utilité des états.

La valeur d'action fournit une perspective détaillée de l'utilité des actions dans ces états, en tenant compte de la politique de l'agent.

3. Dans un état où Pac-Man est à côté d'une pac-gomme, si la fonction de valeur d'état pour cet état est de $V(s) = 25$, quelle serait la valeur d'état mise à jour après avoir mangé la pac-gomme ?

Après avoir mangé la pac-gomme, la récompense immédiate est de +10. La nouvelle valeur d'état serait la somme de la récompense immédiate et de la valeur d'état future : $V(s) = 10 + 25 = 35$.

4. Si Pac-Man peut se déplacer dans un état voisin et la récompense attendue dans est de 10, quelle serait la fonction de valeur d'état mise à jour après avoir pris une certaine action ?

La fonction de valeur d'état mise à jour serait calculée en prenant en compte la récompense immédiate et la valeur d'état future pondérée par les probabilités de transition : $V(s) = 0.8 * 10 + 0.2 * 0 = 8$.

5. Supposons que dans un état où Pac-Man peut se déplacer vers le haut, la récompense immédiate de prendre cette action est de +5. Si la valeur d'état après avoir pris cette action est de $V(s') = 30$, quelle serait la valeur d'action mise à jour pour cette action ?

La valeur d'action mise à jour serait la somme de la récompense immédiate et de la valeur d'état future : $Q(s, a) = 5 + 30 = 35$.

6. Si les valeurs d'action pour un état donné sont $Q(s, haut) = 12$, $Q(s, bas) = 15$, $Q(s, gauche) = 18$ et $Q(s, droite) = 20$, comment pouvez-vous calculer la fonction de valeur d'état pour cet état ?

Pour calculer la fonction de valeur d'état à partir des valeurs d'action, vous pouvez prendre la valeur maximale (20), ou prendre la moyenne des valeurs d'action dans cet état. Dans cet exemple, cela donnerait : $V(s) = (12 + 15 + 18 + 20) / 4 = 16.25$.

7. Comment la table Q est-elle structurée dans le contexte de l'apprentissage par renforcement de Pac-Man, et quelles informations sont stockées dans cette table ?

Etat	Haut	Bas	Gauche	Droite
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
...	0	0	0	0
N	0	0	0	0

Dans ce tableau Q, chaque ligne correspond à un état possible dans le jeu, et chaque colonne représente une action que Pac-Man peut prendre depuis cet état (Haut, Bas, Gauche, Droite). Les valeurs indiquées dans chaque cellule représentent la récompense attendue lorsque Pac-Man prend cette action dans cet état.