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° 5

❖ Exercice 1 - Dilemme du Prisonnier répété

Dans une version répétée du dilemme du prisonnier, deux agents, Alpha et Beta, s'affrontent dans une série de rounds où chacun peut soit coopérer (C) soit trahir (T). Les récompenses sont définies comme suit : coopération mutuelle = 3 points chacun, trahison mutuelle = 1 point chacun, une trahison face à une coopération = 5 points pour le traître et 0 pour le coopérant.

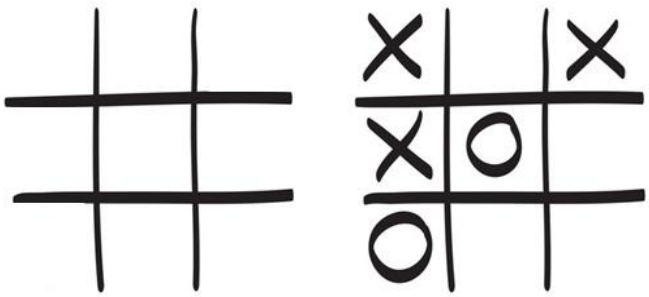
		Joueur Beta			
		Coopérer		Trahir	
Joueur Alpha	Coopérer	α = 3	β = 3	α = 0	β = 5
	Trahir	α = 5	β = 0	α = 1	β = 1

Questions

1. Définir les concepts d'exploration et d'exploitation dans le contexte de ce jeu.
2. Expliquer comment la stratégie ε-Greedy pourrait être utilisée par les joueurs dans ce contexte.
3. Si Alpha utilise un ε de 0.1 et commence à coopérer mais Beta trahit constamment, quelle modification Alpha pourrait-il envisager pour son ε après 10 rounds? Justifiez pourquoi cette modification pourrait être bénéfique.
4. Proposer une stratégie alternative à ε-Greedy pour ce jeu.
5. Quelle serait l'issue probable si les deux agents décidaient d'augmenter simultanément leur ε à 0.5 après 50 rounds? Discutez des implications possibles sur leur stratégie globale et les scores finaux.
6. Quelles sont les conséquences possibles d'une coopération continue versus trahison continue dans la stratégie ε-Greedy ?
7. Comment l'utilisation de la stratégie ε-Greedy dans le dilemme du prisonnier répété peut-elle être vue à travers le prisme de l'éthique? Discutez de l'impact d'une telle stratégie sur la confiance entre les agents.

❖ Exercice 2 – Tic Tac Toe

Dans ce jeu, deux joueurs placent alternativement des symboles (X et O) sur une grille 3x3. Le premier joueur à aligner trois de ses symboles horizontalement, verticalement ou en diagonale gagne la partie. Un agent d'apprentissage par renforcement (AI) joue contre un adversaire qui choisit ses coups de manière aléatoire. L'agent (AI) doit utiliser la stratégie ε-Greedy pour améliorer sa performance.
L'objectif pour l'agent (AI) est d'apprendre à maximiser ses chances de gagner en utilisant les expériences des parties précédentes, tout en s'assurant de continuer à explorer de nouvelles stratégies pour ne pas rater potentiellement de meilleures tactiques.



Questions

1. Quelles sont les actions spécifiques que cet agent peut choisir lors de son tour ?
2. Expliquer comment l'utilisation de la stratégie ε-Greedy permet à l'agent d'équilibrer entre exploration de nouveaux coups et exploitation des coups qui ont le plus souvent mené à la victoire.
3. Proposer une méthode pour ajuster ε au fil du temps. Quels facteurs pourraient influencer cet ajustement?
4. Comment mesurer l'efficacité de la stratégie ε-Greedy dans ce jeu?
5. Proposez une stratégie innovante que l'AI pourrait utiliser après avoir détecté un motif récurrent dans les coups de l'adversaire. Comment cette stratégie utiliserait-elle les principes d'exploration et d'exploitation?
6. Si l'AI commence à perdre des parties, comment devrait-elle ajuster son ε pour rétablir sa compétitivité?
7. Comment l'adaptabilité de l'AI à l'aide de la stratégie ε-Greedy pourrait-elle être testée en changeant les règles du jeu de tic-tac-toe (par exemple, en jouant sur un tableau plus grand)? Quels défis nouveaux cela pourrait-il introduire pour l'AI?

❖ **Solution exercice 1 :**

1. **Exploration** signifie choisir au hasard entre coopérer (C) et trahir (T), sans se baser sur les résultats précédents. Cela permet aux joueurs de tester différentes stratégies et de découvrir de meilleures approches en fonction des réactions de l'autre joueur.
Exploitation consiste à choisir l'action qui a historiquement donné les meilleurs résultats, basé sur les récompenses accumulées jusqu'à présent. Si trahir a souvent conduit à une récompense plus élevée, un joueur exploitant continuerait à trahir.

2. Chaque joueur a une probabilité ε de choisir aléatoirement son action (exploration) et une probabilité 1-ε de choisir l'action qui maximise ses gains historiques (exploitation). Initialement, sans données, les actions peuvent être choisies au hasard, mais avec le temps, les choix se baseront de plus en plus sur les actions passées.

3. Si Alpha commence avec un ε de 0.1 et continue à coopérer pendant que Beta trahit constamment, Alpha pourrait envisager d'augmenter son ε pour explorer plus fréquemment d'autres stratégies, notamment la trahison. Ceci pourrait aider à mieux se protéger contre les pertes conséquentes dues à la coopération unilatérale, tout en cherchant potentiellement à déstabiliser la stratégie constante de Beta.

4. Stratégie Tit-for-Tat : Initialement coopérer, puis répliquer l'action du dernier tour de l'adversaire. C'est une stratégie qui favorise la coopération mais peut être vulnérable si l'autre joueur change souvent de stratégie.

5. En augmentant leur ε à 0.5, les deux agents commenceraient à alterner plus fréquemment entre coopération et trahison. Cela pourrait entraîner un environnement de jeu plus imprévisible et potentiellement moins optimal, car les deux stratégies seraient choisies presque au hasard, ce qui réduirait la possibilité d'apprendre et d'exploiter efficacement les actions de l'autre.

6. Si les deux joueurs continuent de coopérer, ils obtiennent un score stable et élevé (3 points à chaque fois). Cependant, un joueur pourrait être tenté de trahir si le gain immédiat semble plus élevé (5 points), surtout si l'autre joueur continue de coopérer.
Si un joueur commence à trahir régulièrement et l'autre exploite également en trahissant, ils finiront tous les deux par obtenir un score plus bas (1 point chacun).

7. Cette stratégie peut montrer l'importance de la réputation, la confiance à long terme par rapport aux gains à court terme, et comment les stratégies égoïstes peuvent mener à des résultats sous-optimaux pour tous les participants.

❖ **Solution exercice 2 :**

1. L'agent qui joue le rôle du joueur "X" peut choisir de placer un "X" dans l'une des cases libres du plateau. Chaque cas peut être identifié par un indice basé sur sa position.

2. L'utilisation de la stratégie ε-Greedy permet à l'agent de choisir entre placer un O ou un X dans une case spécifique. Avec une probabilité ε, l'agent choisira aléatoirement une case pour explorer de nouveaux coups. Avec une probabilité $1-\epsilon$, il choisira la case qui a historiquement mené à la victoire pour exploiter cette stratégie.

3. L'ajustement de ε au fil du temps pourrait être basé sur la performance de l'agent. Par exemple, si l'agent gagne régulièrement, ε pourrait être réduit pour privilégier l'exploitation des stratégies gagnantes. Si l'agent perd régulièrement, ε pourrait être augmenté pour encourager davantage d'exploration. Les facteurs influençant cet ajustement pourraient inclure le nombre de parties jouées, les résultats des parties précédentes et la stratégie de l'adversaire.

4. L'efficacité de la stratégie ε-Greedy pourrait être mesurée par le taux de victoires de l'agent sur un grand nombre de parties. On pourrait également examiner la diversité des positions choisies par l'agent pour évaluer son équilibre entre exploration et exploitation.

5. Après avoir remarqué que l'adversaire effectue souvent des mouvements similaires, l'agent peut choisir de bloquer les mouvements prévisibles de l'adversaire. En même temps, il peut placer ses propres symboles de manière à créer des opportunités de victoire. Cela lui permet de contrer les stratégies connues de l'adversaire tout en explorant de nouvelles façons de gagner.

6. Si l'agent commence à perdre des parties après une série de victoires, il pourrait augmenter ε pour encourager davantage d'exploration et éviter d'être piégé dans des stratégies inefficaces. Cela permettrait à l'agent de rétablir sa compétitivité en réévaluant ses choix de placer les symboles O et X.

7. L'adaptabilité de l'agent pourrait être testée en changeant les règles du jeu, par exemple en jouant sur un tableau plus grand ou en introduisant des règles supplémentaires pour placer les symboles O et X. Cela introduirait de nouveaux défis pour l'agent, tels que la gestion de plus de combinaisons possibles et la nécessité d'une stratégie plus complexe pour placer les symboles dans des positions stratégiques.