Quiz : PPO et DQL - Concepts et Applications

Partie 1 : Concepts généraux (10 questions)

1. Qu'est-ce que PPO essaie de maximiser?

1. La table Q
2. Les récompenses cumulées attendues
3. Les valeurs d'état
4. Les pénalités

2. Quel est le rôle principal du réseau de politique π_θ?

1. Décider quelle action prendre
2. Évaluer la qualité d'un état
3. Limiter les changements brusques
4. Ajuster les récompenses

3. Quel est le rôle principal du réseau de valeur V_ϕ?

1. Décider quelle action prendre
2. Prédire la qualité d'un état
3. Augmenter les probabilités d'une action
4. Explorer de nouvelles actions

4. Pourquoi PPO utilise-t-il le clipping?

1. Pour réduire la taille des données
2. Pour limiter les changements brusques dans la politique
3. Pour encourager l'exploration
4. Pour optimiser la mémoire

5. Quel concept PPO utilise-t-il pour encourager l'exploration?

1. Clipping
2. Pénalité d'entropie
3. Réduction des valeurs Q
4. Réseaux convolutifs

6. Quelle est la principale différence entre PPO et DQL?

1. PPO utilise une table Q, DQL utilise un réseau de neurones
2. PPO optimise directement une politique, DQL optimise une valeur Q
3. PPO est basé sur des actions discrètes, DQL sur des actions continues
4. DQL favorise l'exploration, PPO favorise l'exploitation

7. Quel est le rôle des récompenses dans PPO?

1. Encourager des actions spécifiques
2. Mettre à jour directement le réseau de valeur
3. Optimiser la mémoire de l'agent
4. Indiquer si une action était bonne ou mauvaise

8. Dans quel type d'environnement PPO est-il particulièrement adapté?

1. Environnements discrets
2. Environnements continus ou complexes
3. Jeux de stratégie au tour par tour
4. Systèmes où la politique est fixe

9. Qu'est-ce que DQL utilise pour stocker les informations sur les actions?

1. Une politique probabiliste
2. Une table Q approximée par un réseau de neurones
3. Une carte des récompenses cumulées
4. Des poids aléatoires

10. Pourquoi PPO utilise-t-il deux réseaux au lieu d'un seul?

1. Pour simplifier l'entraînement
2. Pour séparer les tâches de décision et d'évaluation
3. Pour réduire la taille de l'espace mémoire
4. Pour limiter l'exploration

Partie 2 : Applications et Comparaisons (20 questions)

11. Dans quel type de jeu DQL est-il le plus efficace?

1. Jeux de simulation complexes
2. Jeux vidéo simples avec actions discrètes
3. Jeux de rôle massivement multijoueurs
4. Jeux nécessitant des décisions continues

12. Pourquoi PPO est-il mieux adapté que DQL pour des véhicules autonomes?

1. PPO est plus rapide
2. PPO gère mieux les actions continues
3. PPO utilise une table Q
4. PPO ne nécessite pas de récompenses

13. Quel algorithme serait idéal pour contrôler un bras robotisé dans une usine?

1. DQL
2. PPO
3. TD Learning
4. Monte Carlo

14. Dans quel scénario DQL devient-il inefficace?

1. Quand les actions sont discrètes
2. Quand l'environnement est trop simple
3. Quand l'espace d'action est continu ou très large
4. Quand la politique est déterministe

15. Comment PPO gère-t-il les actions continues?

1. En discrétisant l'espace d'action
2. En optimisant une politique probabiliste
3. En utilisant une table de recherche
4. En calculant directement les valeurs Q

16. Dans une tâche où un robot doit éviter des obstacles et ramasser des objets, quel algorithme choisiriez-vous?

1. PPO
2. DQL
3. TD Learning
4. Monte Carlo

17. Pourquoi PPO est-il préféré dans des simulations industrielles complexes?

1. Il converge plus rapidement
2. Il limite les mises à jour instables de la politique
3. Il est plus facile à implémenter
4. Il fonctionne sans réseau de valeur

18. Dans quel cas DQL est-il supérieur à PPO?

1. Actions continues
2. Environnements très complexes
3. Espaces d'action limités et discrets
4. Exploration agressive

19. Quel algorithme est le plus adapté pour optimiser l'énergie dans un data center?

1. PPO
2. DQL
3. Q-Learning classique
4. Monte Carlo

20. Quel est le rôle de la pénalité d'entropie dans PPO?

1. Encourager des actions imprévisibles
2. Réduire les récompenses excessives
3. Limiter les changements de la politique
4. Stabiliser le réseau de valeur

Questions avancées pour comprendre les différences pratiques

21. Dans quel cas DQL ne peut-il pas gérer les données efficacement?

1. Environnements continus
2. Jeux avec des récompenses immédiates
3. Espaces d'actions petits et fixes
4. Quand les trajectoires sont prévisibles

22. Quel est l'inconvénient majeur d'utiliser une table Q dans DQL?

1. Elle ne peut pas généraliser aux nouvelles situations
2. Elle ne fonctionne qu'avec des réseaux convolutifs
3. Elle nécessite une politique fixe
4. Elle est instable dans des environnements simples

23. Comment PPO limite-t-il les changements brutaux dans la politique?

1. Avec des réseaux convolutifs
2. Grâce à la fonction de clipping
3. En optimisant directement $Q(s, a)$
4. En évitant les mises à jour des probabilités

24. Pourquoi DQL nécessite-t-il un mécanisme d'exploration ϵ-greedy?

1. Pour limiter les erreurs de calcul
2. Pour éviter que l'agent se fixe sur une seule stratégie trop tôt
3. Pour accélérer la convergence
4. Pour stabiliser la table Q

25. Quel type de tâche PPO est-il incapable de gérer efficacement?

1. Actions discrètes simples
2. Actions continues complexes
3. Environnements avec de longues récompenses différées
4. Jeux à très faible complexité

26. Quel algorithme choisiriez-vous pour un jeu vidéo avec des actions prédéfinies (aller à gauche, droite, sauter)?

1. PPO
2. DQL
3. TD Learning
4. Monte Carlo

27. Quel est l'avantage d'utiliser deux réseaux dans PPO?

1. Réduire la mémoire requise
2. Séparer les tâches de décision et d'évaluation
3. Rendre les actions déterministes
4. Réduire le temps d'entraînement

28. Quel algorithme est idéal pour contrôler un drone volant dans toutes les directions?

1. PPO
2. DQL
3. TD Learning
4. Monte Carlo

29. Pourquoi PPO est souvent préféré dans des environnements incertains ?

1. Il est plus rapide à implémenter
2. Il évite les mises à jour instables grâce au clipping
3. Il n'utilise pas de récompenses cumulées
4. Il fonctionne avec une politique fixe

30. Dans une application de trading, pourquoi PPO est-il plus adapté?

1. Il gère des actions discrètes limitées
2. Il optimise une politique continue, comme les quantités à acheter/vendre
3. Il ne nécessite pas de réseaux de neurones
4. Il réduit les pertes sans exploration