Concept de DQN (Deep Q-Network)

Introduction

Le concept de **DQN (Deep Q-Network)** est une extension de l'algorithme de **Q-Learning**, adaptée pour résoudre des problèmes complexes où les espaces d'états sont trop grands pour être représentés explicitement. DQN utilise des réseaux de neurones profonds pour approximer la fonction de valeur d'action $Q(s, a)$, qui est essentielle dans le cadre des algorithmes d'apprentissage par renforcement.

Le concept des deux réseaux dans DQN

DQN introduit l'utilisation de **deux réseaux neuronaux distincts** pour stabiliser l'apprentissage : le **réseau principal (ou réseau en ligne)** et le **réseau cible**.

1. Réseau principal (Online Network)

- Ce réseau est utilisé pour **choisir les actions** en fonction de l'état actuel s.
- Pendant l'entraînement, il approxime la fonction $Q(s, a; \theta)$, où θ représente les poids du réseau.
- C'est le réseau qui est mis à jour directement à chaque itération en minimisant une perte (fonction de coût) définie par la différence entre la valeur Q prédite et la valeur Q cible.

2. Réseau cible (Target Network)

- Ce réseau est utilisé pour calculer les **valeurs cibles** lors de l'entraînement du réseau principal.
- Ses poids (θ^-) sont mis à jour **moins fréquemment** (par exemple, toutes les N itérations), ce qui aide à stabiliser l'apprentissage en fournissant une référence relativement fixe pour les mises à jour.
- La fonction cible est définie comme :

$$y = r + \gamma \max_{a'} Q(s', a'; \theta^-)$$

où :

  - r est la récompense immédiate,
  - γ est le facteur d'actualisation,
  - s' est l'état suivant,
  - θ^- représente les poids du réseau cible.

Pourquoi deux réseaux sont-ils nécessaires ?

1. **Stabilisation de l'apprentissage** :
   - Si le même réseau est utilisé pour calculer à la fois les valeurs cibles et les valeurs prédites, les poids du réseau changeraient constamment, rendant l'apprentissage instable.
   - Le réseau cible fournit une "ancre" pour les mises à jour.
2. **Réduction des oscillations et des divergences** :

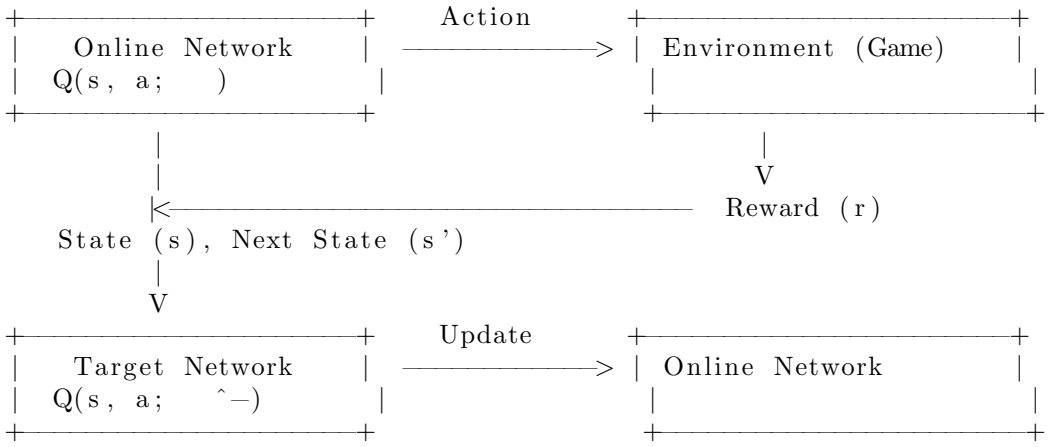
- En maintenant une référence plus stable avec le réseau cible, le modèle converge plus facilement.

3. **Amélioration de l'efficacité** :
   - Cela permet à DQN de résoudre des problèmes avec des espaces d'états énormes, comme ceux rencontrés dans les environnements de jeux Atari.

Fonctionnement global de DQN

1. **Initialisation** :
   - Deux réseaux neuronaux sont créés : le réseau principal et le réseau cible.
   - Les poids du réseau cible sont initialisés à partir des poids du réseau principal.
2. **Entraînement** :
   - Le réseau principal est entraîné à chaque étape à partir d'un ensemble de transitions (s, a, r, s') stockées dans une mémoire de répétition d'expérience (**Experience Replay**).
   - Le réseau cible est mis à jour périodiquement en copiant les poids du réseau principal.
3. **Prise de décision** :
   - Lors de l'interaction avec l'environnement, le réseau principal est utilisé pour choisir les actions.

Résumé



En résumé, les deux réseaux de DQN permettent d'approximer la fonction $Q(s, a)$ de manière stable, rendant l'apprentissage plus robuste et efficace.