Deep Reinforcement Learning

L'Agent DQN de Nouvelle Génération

au Service de la Finance

Améliorations du DQN : Double DQN, Dueling DQN, Prioritized Experience Replay

Application au Trading Financier via TradingEnv (Gymnasium)

Le Défi du DQN Classique

Un Optimisme Dangereux

Rappel : Principe du DQN

Un réseau de neurones estime la valeur future des actions (**Q-values**) : Q(s,a) = valeur attendue de prendre l'action a dans l'état s.

Le Problème Critique :

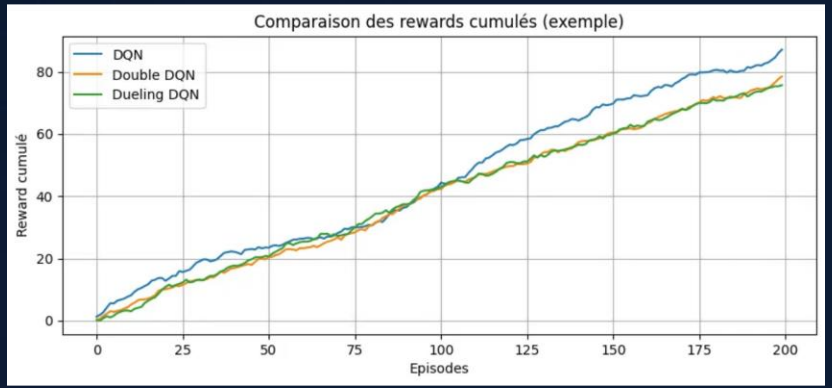
Le DQN standard surestime systématiquement les Q-values. Le même réseau est utilisé pour **sélectionner** l'action future ET pour **évaluer** sa valeur, créant un biais optimiste.

- Conséquence

Instabilité de l'apprentissage et politiques sous-optimales

Réponse

Trois piliers d'amélioration pour un agent stable et performant



Double DQN

La Stabilité par le Double Contrôle

Mécanisme de Séparation des Rôles

1. **Sélection :** Le réseau **Online** choisit l'action future $a^* = \text{argmax}\ Q_{online}(s', a)$
2. **Évaluation :** Le réseau **Target** évalue cette action $Q_{target}(s', a^*)$
3. **Cible :** $\text{Target} = r + \gamma \times Q_{target}(s', a^*) \times (1 - \text{done})$

Formule Clé :

$a^* \leftarrow \text{argmax}_a\ Q_{online}(s', a)$

$Q_{target} \leftarrow r + \gamma \times Q_{target}(s', a^*)$

$\text{Loss} \leftarrow \text{MSE}(Q_{online}(s,a), Q_{target})$

Avantage

Réduction drastique du biais de surestimation

Impact

Apprentissage plus robuste et fiable

Dueling DQN

Une Architecture qui Comprend l'État

Décomposition de la Q-Value

$$Q(s,a) = V(s) + (A(s,a) - \text{mean}(A(s,a')))$$

V(s) — Value Function

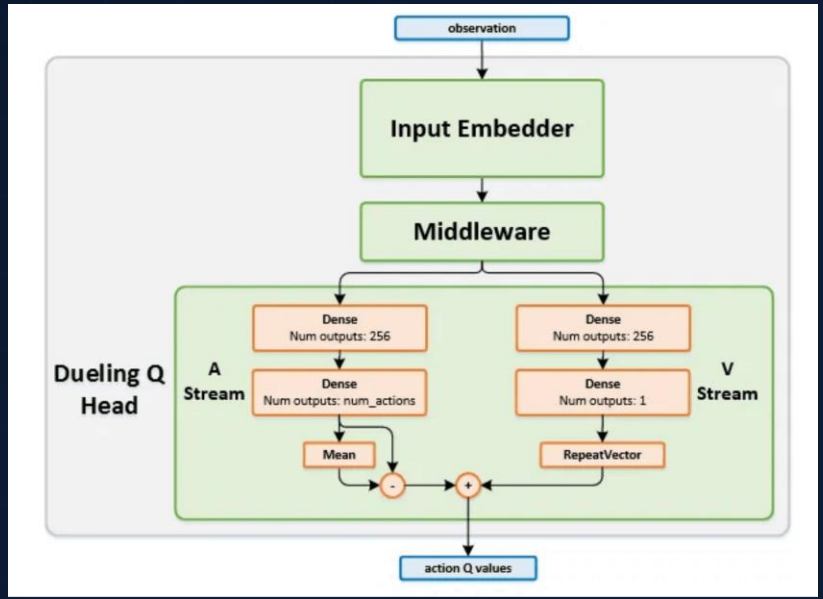
Valeur intrinsèque de l'état, indépendante de l'action. Répond à : "À quel point cet état est-il bon ?"

A(s,a) — Advantage Function

Avantage relatif de chaque action. Répond à : "Quelle action est meilleure que les autres ?"

Avantage Clé

L'agent apprend **séparément** la valeur des états et l'avantage des actions, conduisant à une



Prioritized Experience Replay

Concentrer l'Apprentissage sur les Leçons les Plus Riches

Le Problème : Échantillonnage Uniforme

Approche Standard

Échantillonnage aléatoire uniforme du buffer. Toutes les transitions ont la même probabilité.

Approche PER

Échantillonnage intelligent basé sur l'**erreur TD**. Transitions "surprenantes" rejouées plus souvent.

Mécanisme PER

Priorité : $p_i = |\delta_i|^{\alpha}$

Probabilité : $P(i) = p_i / \Sigma p_j$

IS Weights : $w_i = (1 / (N \times P(i)))^{\beta}$

Annealing : β augmente de β_{start} à 1.0

Paramètres Clés

α (Alpha)

Contrôle l'importance de la priorité (0 = uniforme, 1 = full)

β (Beta)

Correction du biais (~0.4 → 1.0)

Impact

Efficacité : Apprentissage plus rapide sur données informatives

L'Agent Intégré

Le DoubleDuelingDQNAgent : Triple-A

DDQN - Robustesse

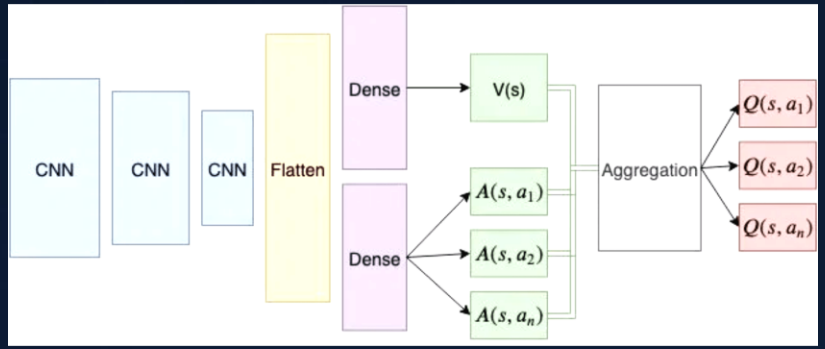
Sépare la sélection et l'évaluation des actions pour éliminer le biais de surestimation. Assure la **stabilité** de l'apprentissage.

Dueling - Efficacité

Décompose la Q-value en Value V(s) et Advantage A(s,a). Améliore la **compréhension** des états et la convergence.

PER - Focus

Échantillonne intelligemment selon le TD-error. Garantit un **apprentissage rapide** et efficace.



Conclusion : L'agent **DoubleDuelingDQNAgent** combine les trois techniques pour créer un modèle de décision **robuste, efficace et ciblé**, prêt à affronter la complexité des environnements réels comme le trading financier.

Application Concrète

Le TradingEnv (Gymnasium)

Environnement : Le **TradingEnv** modélise un marché financier réaliste en utilisant le standard **Gymnasium**, permettant à l'agent DRL de s'entraîner à prendre des décisions de trading optimales.

Espace d'Observation

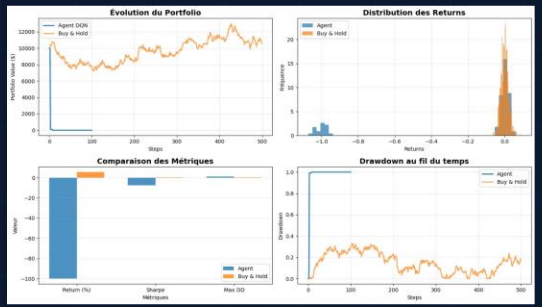
- Prix historiques
- Indicateurs techniques
- Volatilité du marché
- État du portefeuille

Espace d'Action

- **0 : Acheter**
- **1 : Vendre**
- **2 : Conserver**

Fonction de Récompense

- Changement de portefeuille
- Pénalité pour transactions
- Bonus pour rendements



Performance de l'agent DRL vs stratégie Buy & Hold : évolution du portefeuille, distribution des rendements, métriques de performance, et drawdown

Objectif de l'Agent

Maximiser la richesse du portefeuille à long terme en apprenant une politique optimale de trading. L'agent utilise les améliorations du DQN pour naviguer efficacement dans cet environnement complexe et incertain.

Conclusion

Impact Majeur des Améliorations du DQN

Les améliorations du DQN — **Double DQN, Dueling DQN et PER** — transforment l'algorithme en un outil **fiable, rapide et intelligent** pour l'automatisation des décisions dans des environnements complexes.

DDQN

Stabilité par séparation des rôles de sélection et d'évaluation

Dueling

Efficacité architecturale et convergence rapide

PER

Focus sur les données les plus informatives

Perspectives

Le Futur de la Décision Financière Automatisée

Applications Réelles

Ces techniques sont essentielles pour tout problème de DRL où la stabilité, la vitesse de convergence et la généralisation sont critiques : **robotique, jeux complexes, finance**.

Prochaines Étapes

Optimisation des hyperparamètres, intégration de **données financières réelles**, comparaison de performance avec les stratégies traditionnelles.

Innovation

Le DRL ouvre la voie à la découverte de **stratégies de trading complexes** que l'humain ou les modèles traditionnels ne pourraient pas identifier. C'est une révolution dans l'automatisation financière.

MERCI