Cours DQN – Section 3.1

3.1 – Replay Memory : revisiter l'expérience passée

L'une des innovations majeures du DQN est l'introduction d'une mémoire de répétition, appelée **Replay Memory**. Cette technique résout plusieurs problèmes liés à l'apprentissage instable et inefficace dans les environnements complexes.

Problème : corrélation temporelle entre les expériences

Dans un apprentissage séquentiel classique :

— L'agent apprend en traitant les expériences une à une, dans l'ordre où elles sont vécues.

— Les données successives sont **corrélées**, car elles proviennent d'états voisins dans le temps.

Conséquences :

— L'apprentissage est biaisé.

— Les gradients deviennent instables.

— Le réseau peut « oublier » des expériences utiles ou surapprendre des séquences spécifiques.

Solution : stocker les expériences dans une mémoire

À chaque interaction, l'agent enregistre son expérience dans une mémoire :

$$\text{expérience} = (s_t, a_t, r_t, s_{t+1}, \text{done})$$

La **Replay Memory** est typiquement une file circulaire de grande capacité (ex. : 1 million d'expériences).

Principe d'apprentissage par relecture

Plutôt que d'apprendre uniquement à partir de l'expérience la plus récente, le réseau est entraîné à partir d'un **mini-batch d'expériences** tirées aléatoirement dans la mémoire.

Avantages :

— Décorrélation temporelle : les expériences sont mélangées.

— Réutilisation efficace des données : on apprend plusieurs fois à partir d'une même expérience.

— Apprentissage plus stable et plus rapide.

Étapes concrètes

1. L'agent joue, collecte des expériences (s, a, r, s') à chaque étape.
2. Chaque expérience est stockée dans la mémoire.
3. À chaque mise à jour, un mini-lot de N expériences est tiré au hasard.
4. Le réseau de neurones est entraîné sur ce mini-lot.

Illustration pédagogique

> C'est comme un élève qui garde un carnet de notes d'erreurs passées, et qui révise régulièrement en tirant des fiches au hasard pour se réentraîner sur des erreurs anciennes.

Conclusion

La Replay Memory est un composant essentiel du DQN. Elle permet :
— Une meilleure efficacité d'apprentissage.
— Une robustesse accrue face à l'instabilité des données séquentielles.
— Une réutilisation intelligente de l'information collectée.

Dans la prochaine section, nous verrons une autre innovation tout aussi importante : le **Target Network**, qui stabilise davantage l'apprentissage.