Le Robot Joueur de Soccer

1 Ton robot joueur de soccer : Le réseau de politique π_θ

Imagine que tu as un **robot joueur de soccer** qui doit apprendre à jouer comme un pro. Le réseau de politique, c'est **son cerveau stratégique**. Il regarde ce qui se passe sur le terrain et décide **quoi faire**, par exemple:

"Dois-je passer? Dribbler? Tirer?"

Comment ça marche dans un match de soccer ?

1. **Entrée du réseau de politique :** Le robot observe ce qui se passe sur le terrain, comme:
   - Où est le ballon?
   - Où sont ses coéquipiers?
   - Où est le but adverse?
   - Où sont les défenseurs?
2. **Traitement par le réseau :** Le réseau réfléchit comme un stratège, par exemple :
   - "Si je suis près du but et sans défenseur devant, je devrais tirer."
   - "Si un coéquipier est mieux placé, je devrais passer."
3. **Sortie du réseau :** Le réseau donne les **probabilités pour chaque action** :
   - Passer : 50 %.
   - Dribbler : 30 %.
   - Tirer : 20 %.

   Le robot choisit ensuite l'action avec la probabilité la plus élevée (par exemple, *"passer"*).

2 Ton coach personnel : Le réseau de valeur V_ϕ

Le réseau de valeur agit comme un **coach qui observe le match** depuis les tribunes. Il ne dit pas directement quoi faire, mais il évalue la situation actuelle:

"Est-ce que ton équipe est dans une bonne position ou non?"

Comment ça marche dans un match de soccer ?

1. **Entrée du réseau de valeur :** Comme le réseau de politique, il observe:
   - Où est le ballon?
   - Où sont les joueurs?
   - Est-ce que le robot est proche du but adverse ou en défense?
2. **Traitement par le réseau :** Le coach réfléchit à la situation globale, par exemple :
   - "Si le robot est dans la moitié adverse avec un coéquipier bien placé, c'est une bonne position."
   - "Si le ballon est proche de notre but, ce n'est pas idéal."
3. **Sortie du réseau :** Le réseau attribue une **évaluation numérique de la situation** :
   - "+10 points" si l'équipe est proche de marquer.
   - "-5 points" si l'équipe est en danger.

Ce réseau aide le robot à comprendre **si ses actions sont en train de l'amener vers une victoire ou un échec.**

3 Les deux réseaux travaillent ensemble

Sur le terrain, les deux réseaux fonctionnent en synergie :

- **Réseau de politique :** Il décide **quoi faire immédiatement**. *"Passe le ballon à ton coéquipier!"*
- **Réseau de valeur :** Il évalue la situation globale. *"C'est une bonne situation, continue comme ça!"*

4 Comment ces réseaux apprennent pendant le match ?

Au début, le robot est un **débutant total**. Il fait n'importe quoi sur le terrain :

- Il tire même quand il est à 50 mètres du but.
- Il dribble dans ses propres cages.

Mais à chaque match, il **apprend de ses erreurs** grâce à PPO.

Étapes d'apprentissage

1. **Collecter des données :** Pendant le match, le robot note tout ce qu'il a fait :
   - "J'ai dribblé ici, j'ai passé là-bas."
   - "J'ai marqué un but, donc cette action était bonne."
   - "J'ai perdu le ballon, donc cette action était mauvaise."
2. **Améliorer le réseau de politique :**
   - Si le robot voit qu'un tir a mené à un but, le réseau de politique apprend à **augmenter la probabilité de tirer dans cette situation**.
   - Si un tir a été bloqué inutilement, le réseau apprend à **réduire la probabilité de tirer dans cette situation**.
3. **Améliorer le réseau de valeur :**
   - Si le réseau de valeur a dit *"Cet état vaut +10 points"*, mais que le robot finit par perdre, il comprend qu'il a **surestimé la situation**.
   - Il ajuste ses prédictions pour être plus précis à l'avenir.

5 Un exemple concret pendant un match

Supposons que le robot est dans une situation où:

- Il est à 10 mètres du but.
- Son coéquipier est mieux placé pour marquer.
- Le gardien adverse est bien positionné.

Ce que fait le réseau de politique π_θ :

- **Options :**
  - Passer? 70 % de chances de réussite, car le coéquipier est bien placé.
  - Tirer? 30 % de chances de réussite, car le gardien peut arrêter.
- **Décision :** Passe le ballon.

Ce que fait le réseau de valeur V_ϕ :

- **Évaluation :** "L'état est favorable, car on est proche du but."
- **Note :** "+15 points."

Le robot passe le ballon, son coéquipier marque, et tout le monde est content.

6 Résumé avec le soccer

1. **Réseau de politique π_θ :** Décide **quelle action prendre** (passer, tirer, dribbler). C'est le **stratège du robot.**

2. **Réseau de valeur V_ϕ :** Évalue **à quel point la situation actuelle est bonne ou mauvaise**. C'est le **coach du robot.**

3. **Ils travaillent ensemble :** Le réseau de politique agit, et le réseau de valeur guide en arrière-plan.

4. **Ils apprennent en jouant :** Plus le robot joue, plus ses décisions deviennent intelligentes et ses évaluations précises.

Avec ces deux réseaux, ton robot peut passer de débutant maladroit à **joueur de classe mondiale**.