Atelier N°9 : Intrinsic Motivation (synthèse)

Arthur YVARS, Danyl-Rayane ALLOUACHE, Khalissa RHOULAM

1. Motivation Intrinsèque vs. Extrinsèque

1.1 Le Problème : Les Limites de l'Extrinsèque

L'Apprentissage par Renforcement (RL) classique, basé sur une **motivation extrinsèque** (récompenses `r_ext` définies par un programmeur), peine dans les environnements où les récompenses sont **rares** (*sparse rewards*) ou retardées. L'agent, manquant de signaux clairs, explore peu efficacement (ex: *Montezuma's Revenge*) et apprend difficilement, résultant en compétences peu généralisables.

1.2 La Solution : Le Concept de Motivation Intrinsèque

Inspirée de la psychologie, la **Motivation Intrinsèque (MI)** permet à l'agent de générer sa propre récompense interne (`r_int`) pour satisfaire sa "curiosité". Le calcul de la récompense se fait par la somme pondérée suivante : `r_total = β * r_ext + α * r_int`. La MI transforme l'**exploration** aléatoire en une recherche active de nouveauté ou d'apprentissage, permettant d'acquérir des compétences de base de manière autonome.

2. Les Mécanismes de la Curiosité Artificielle

On identifie 3 types de curiosité, soit 3 manières de calculer `r_int`:

2.1 Approches Basées sur la Surprise (Erreur de Prédiction)

- **Principe :** `r_int` proportionnel à l'erreur de prédiction du modèle interne (*forward model*). Plus il se trompe, plus il est récompensé. Cela l'encourage à explorer des zones où son modèle du monde est imparfait.
- **Algorithme :** ICM.
- **Limite :** Attirance pour le bruit stochastique (*Noisy TV Problem*).

2.2 Approches Basées sur la Nouveauté (Densité d'États / Connaissance)

- **Principe :** `r_int` inversement proportionnel à la fréquence/densité estimée de l'état visité. L'agent est incité à visiter des états peu rencontrés.
- **Algorithme :** RND.
- **Avantage :** Robuste au bruit, large couverture des états.

2.3 Approches Basées sur l'Apprentissage de Compétences (*Skill Learning*)

- **Principe :** Apprentissage de politiques (*skills*) distinctes (`z`). `r_int` récompense la capacité de l'agent à produire des comportements variés et reconnaissables pour chaque `z`, souvent en maximisant l'Information Mutuelle entre `z` et les états/trajectoires résultants.
- **Algorithme :** DIAYN.
- **Avantage :** Exploration structurée, compétences réutilisables

.

3. L'Évolution vers une Curiosité Adaptative et Régulée

3.1 L'approche IAC d'Oudeyer et al. (2007)

- **Principe :** L'agent maximise son **progrès d'apprentissage** (*learning progress*) en choisissant les actions menant aux régions de l'espace où son erreur de prédiction diminue le plus vite. Cela le concentre sur la **Zone Proximale de Développement** (Vygotsky) : *la zone où une tâche est un défi stimulant mais réalisable, permettant l'apprentissage le plus efficace.*

3.2 L'approche CIM de Zheng et al. (2024)

- **Principe :** Régulation de la curiosité. **Ajustement dynamique des poids** α (curiosité) et β (tâche externe) selon le progrès sur r_ext. Apprentissage contraint de **compétences distinctes** (grâce à la contrainte d'alignement L_a, qui maximise la reconnaissabilité des *skills*) et **dynamiques** (grâce à la maximisation de $H(\varphi(s)|z)$, qui pousse chaque *skill* à explorer largement plutôt qu'à rester statique).
- **Résultats :** Apprentissage plus efficace, compétences plus riches.
- **Interprétation :** Forme de **métacognition artificielle** (réflexion sur son propre processus apprentissage).

4. Défis Actuels et Perspectives du Domaine

4.1 Limites des MI classiques et Actuelles

- Sparsité de r_int dans certains cas complexes.
- Exploration à long terme : **Détachement** (oubli) et **Dérive** (instabilité).
- Solutions : Exploration hiérarchique, basée sur les frontières.

4.2 Axes de Recherche Futurs

- Curiosité hiérarchique et abstraite.
- Gestion de la mémoire et de la curiosité à long terme.
- **Alignement et Contrôle Éthique :** Défi majeur pour les IA autonomes.

5. Synthèse des Débats

5.1 Débat 1 : Dangerosité de l'IA Curieuse

Conclusion du débat : La discussion a mis en balance les risques liés à l'imprédictibilité d'une IA auto-motivée (décisions inattendues, absence d'éthique innée) et la possibilité d'encadrer cette curiosité (confinement, limitation des actions), ainsi que le potentiel unique de découverte qu'elle offre. Il en ressort que la dangerosité est contextuelle et que des mécanismes de contrôle sont nécessaires mais possibles.

5.2 Débat 2 : Utilité vs. Sécurité (Assistants Personnels)

Conclusion du débat : Face au dilemme entre l'utilité maximale d'un assistant proactif (nécessitant la curiosité) et les risques associés (actions dangereuses, vie privée), le consensus a rejeté l'idée d'un sacrifice total de la sécurité. Des solutions intermédiaires, bornant l'autonomie de l'IA (limites d'action, validation humaine), semblent préférables pour concilier les deux aspects.