IAC-9 : Intrinsic Motivation

UE IA-IAC (INF2346M) - Artificial Intelligence and Cognition - Marie Lefevre

Arthur YVARS
Danyl-Rayane ALLOUACHE
Khalissa RHOULAM

“Et si une IA pouvait apprendre juste par curiosité ?”

Comment apprend une IA ? Deux approches de la récompense

IA Classique (Motivation Extrinsèque)

L'agent reçoit une récompense claire et externe pour avoir atteint un objectif prédéfini.

🏆 "Bravo, tu as gagné !"

IA Curieuse (Motivation Intrinsèque)

L'agent se récompense lui-même pour avoir découvert une nouvelle information ou maîtrisé une nouvelle compétence.

🧠 "Wow, j'ai découvert quelque chose !"

Motivation Extrinsèque vs Intrinsèque

Motivation Extrinsèque

Agir en vue d'obtenir une récompense ou d'éviter une punition provenant de l'environnement externe. **C'est une incitation matérielle ou sociale.**

- Exemple IA : Un robot reçoit +1 point s'il atteint un objectif défini par un programmeur.
- Exemple humain : Recevoir un salaire à la fin du mois pour un travail effectué.

Motivation Intrinsèque

Agir par intérêt personnel, plaisir ou sentiment de satisfaction. **L'activité est sa propre récompense.**

- Exemple IA : Un robot explore une nouvelle zone d'une pièce juste pour voir ce qu'il y a de nouveau.
- Exemple humain : Un enfant qui empile des cubes par pur plaisir de créer et d'expérimenter.

Apprentissage par soi-même

Pourquoi la motivation intrinsèque en IA ?

Dans certains environnements complexes, les récompenses externes ne suffisent plus. C'est là que l'IA curieuse devient indispensable.

Peu ou Pas de Récompenses

Dans un environnement inconnu (comme un robot sur Mars), comment savoir ce qui est bien ou mal ?

Manque de Créativité et d'Exploration

L'IA extrinsèque ne fait que ce qu'on lui demande de faire. Elle ne découvre jamais de nouvelles stratégies imprévues, limitant son potentiel d'innovation.

Dépendance et Passivité

Sans la promesse d'une récompense, l'agent reste passif et ne prend pas d'initiatives. Si le système de notation s'arrête, l'apprentissage s'arrête.

Avantage de la curiosité

En injectant de la curiosité (une forme de motivation intrinsèque), l'IA surmonte les limites des systèmes basés uniquement sur la récompense externe.

1

Exploration Sans Objectif

L'IA explore activement son environnement même en l'absence de récompense directe, privilégiant les zones nouvelles ou imprévisibles.

2

Créativité

Elle trouve des solutions qu'on n'avait pas prévues.
Comme AlphaGo qui a inventé des coups de Go que les humains ne connaissaient pas.

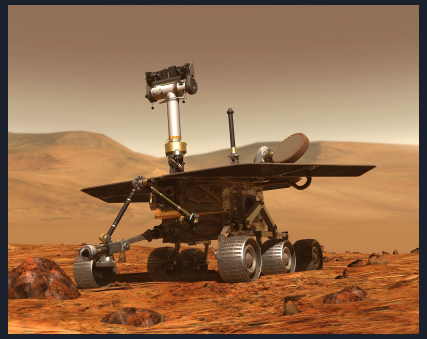
3

Adaptation

L'IA s'adapte à des situations nouvelles,parce qu'elle a envie de comprendre.

> Exemple concret : Un robot apprend à ouvrir une porte sans savoir que c'est utile. Cette compétence de "pousser" ou "manipuler" servira plus tard pour des tâches complexes comme construire ou secourir.

Est-ce utile pour toutes les IA ?



La motivation intrinsèque est intéressante, mais elle n'est pas nécessaire pour tous les systèmes d'Intelligence Artificielle.

Cas où c'est Utile

Robots Explorateurs

Dans des environnements réels où les récompenses sont rares (exploration sous-marine, planétaire, entrepôts non cartographiés).Le robot doit apprendre par lui même ce qui est intéressant.

Agents de Jeux Avancés

Pour découvrir des stratégies non conventionnelles que les programmeurs n'ont pas prévues.

Assistants Personnels

Pour apprendre les habitudes et les préférences de l'utilisateur de manière proactive, sans nécessiter d'instruction explicite constante.

Cas où c'est Inutile

- Tâches Simples et Répétitives
  Un filtre anti-spam n'a pas besoin d'être curieux. Il doit simplement suivre des règles d'identification très claires et précises.
- Classifieurs
  Reconnaître des objets dans des photos (ex. : identifier des chats) ne demande que de la puissance de calcul et des jeux de données étiquetés, pas de l'initiative.
- Systèmes Déterministes
  Les systèmes qui doivent garantir le même résultat à chaque fois (contrôle qualité, processus industriels rigides) bénéficient peu de l'imprévisibilité de la curiosité.

IAC en action

L'IAC (Intelligent Adaptive Curiosity) permet de développer ses compétences de manière autonome, en s'inspirant du développement cognitif des enfants

Étape 1 : Phase Exploration

Au début le robot ne sait rien. Il agit au hasard: il tourne la tête dans toutes les directions, essaie de mordre, sans but précis. Puis il enregistre chaque expérience sous forme d'un vecteur sensorimoteur.

Étape 2 : Création de régions et d'experts

Division de l'espace sensorimoteur en régions en fonction des similitudes entre les expériences. Puis pour chaque région le robot crée un expert (réseau de neurones) pour apprendre à prédire ce qui va se passer si le robot répète une action similaire dans cette région.

Étape 3 : Choix prochaine action

Pour chaque région, le robot calcule son progrès d'apprentissage, puis il choisit l'action qui maximise son progrès d'apprentissage dans une région donnée, et quand une région est maîtrisé il passe à des actions plus complexes qui offrent un nouveau progrès d'apprentissage.

An Information-Theoretic Perspective on Intrinsic Motivation in Reinforcement Learning: A Survey

De Arthur Aubret, Laetitia Matignon et Salima Hassas, dans *Entropy* en 2023

Univ Lyon, UCBL, CNRS, INSA Lyon, LIRIS, UMR5205, 69622 Villeurbanne, France

Le Reinforcement Learning est un domaine en pleine expansion, encore plus depuis l'apparition du Deep Reinforcement Learning.

Le principe : un agent apprend par essai-erreur à maximiser une récompense issue de ses actions dans un environnement.

Ces récompenses sont généralement extrinsèques

↳Est-il intéressant d'y utiliser les récompenses intrinsèques ? Si oui pourquoi et qu'apporte t elles ?

Lien de l'article : https://www.mdpi.com/1099-4300/25/2/327

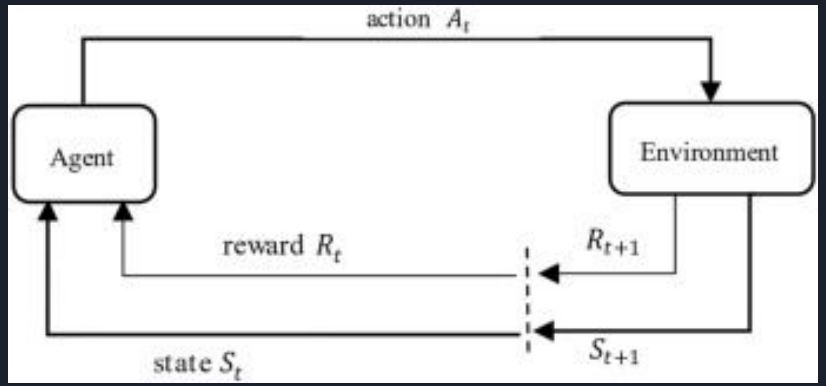
Contexte : limites du Reinforcement Learning et DRL

En RL mais aussi en DRL, un agent apprend à maximiser une récompense extrinsèque.

↳ Problème : si les récompenses sont rares ou retardées, l'apprentissage échoue ou quand les comportements appris ne sont pas réutilisables entre différentes tâches.

↳ Les agents peinent à explorer leur environnement

↳ Pas de généralisation des compétences apprises



Model of RL

Apport de la Motivation Intrinsèque

Dans le contexte du reinforcement learning , la motivation intrinsèque permettrait à un agent :

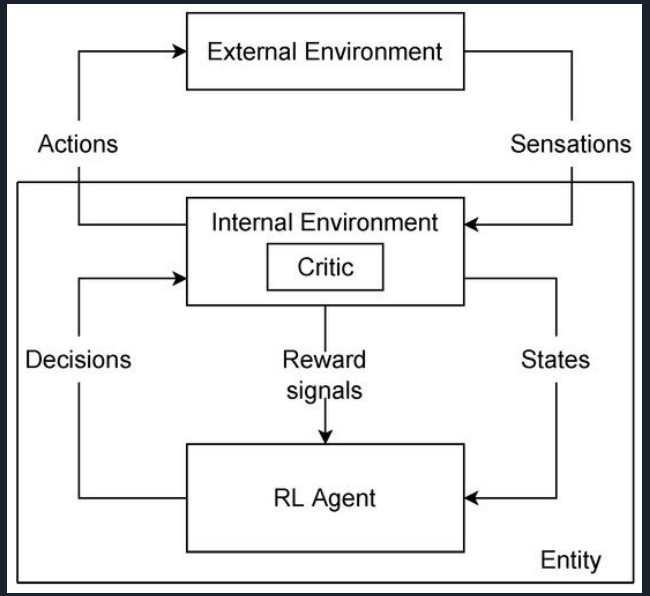
↳ d'explorer son environnement même sans récompense externe immédiate
↳ d'acquérir des compétences générales (curiosité)
↳ de mieux apprendre à long terme en réutilisant ses compétences pour explorer

Les auteurs propose un modèle séparé en deux parties :

Partie externe : représente le monde réel et les tâches à accomplir.
Partie interne : représente les processus cognitifs de l'agent, où sont calculées les récompenses intrinsèques et la valeur des actions.

Le critic calcule les récompenses en combinant les récompenses intrinsèques et extrinsèques selon une pondération :

r = α*rint+ β*rextr, où α et β contrôlent l'importance de chaque type de récompense.



Model of RL integrating IM

Classification des motivations intrinsèques

La Surprise

La surprise correspond à l'écart entre ce que l'agent prévoit et ce qu'il observe réellement.

↳ Prediction Error : mesure à quel point l'agent s'est trompé et le récompense en conséquence.

↳ Learning Progress : l'agent est récompensé par l'amélioration de sa prédiction.

↳ Information Gain : l'agent est récompensé si sa croyance sur sur une prédiction à beaucoup changé (calculé par la KL-divergence).

La Novelty

La novelty mesure à quel point un état est différent de ceux déjà rencontrés.

↳ Information Gain sur un modèle de densité : l'agent apprend un modèle de densité ρ sur les états et cherche à explorer ceux à faible probabilité.

↳ Comptage implicite : mesure la rareté d'un état via le modèle de densité appris (adaptation aux modèles continus).

Le Skill Learning

L'objectif du skill-learning est de permettre à un agent d'acquérir des compétences réutilisables.

Méthodes classiques :

↳ HAC : apprentissage de plusieurs niveaux de politiques, chaque niveau fixant un but pour le niveau inférieur.

↳ HIRO : le but est défini comme la différence entre l'état initial et final.

Perspectives du domaine

Deux difficultés majeures :

Sparsité de la récompense intrinsèque :

- L'agent ne perçoit pas toujours le lien entre ses actions et la récompense (ex. : appuyer sur un bouton dont l'effet est visible bien plus tard).

- Cela empêche les modèles de prédiction de fournir un signal utile.

Montezuma's Revenge → il faut éviter d'utiliser une clé trop tôt, afin de pouvoir s'en servir plus tard.

Exemple d'un niveau de Montezuma's Revenge

Perspectives du domaine

Deux difficultés majeures :

Detachment et Derailment :

- Detachment : l'agent oublie des zones éloignées mais intéressantes à cause du *catastrophic forgetting*.

- Derailment : l'agent échoue à suivre des séquences d'actions précises pour atteindre des zones lointaines à cause de la stochasticité locale.

Pitfall! → on obtient une récompense uniquement après plusieurs salles traversées

Exemple d'un niveau de Pitfall!

Perspectives du domaine

Solutions :

Exploration hiérarchique (pour planifier à long terme)

Exploration basée sur les frontières pour atteindre des zones éloignées avant d'y explorer localement.

Trois axes principaux :

Type de motivation intrinsèque	Objectif principal	Rôle
Surprise	Maximiser l'information entre les modèles internes et l'environnement	Aide à explorer et à planifier
Novelty	Maximiser l'entropie des représentations (diversité des états visités)	Permet d'explorer et de construire de bonnes représentations
Skill-learning	Maximiser l'information entre un but et la trajectoire associée	Permet d'apprendre des compétences hiérarchiques

Cependant, encore limité dans des environnements complexes.

Le Problème : Une Curiosité Sans Limites

- Les approches classiques ont des limites déjà évoquées :
  - **Oubli catastrophique** (*detachment*) : l'agent oublie des zones intéressantes.
  - **Dérive** (*derailment*) : l'agent est instable et se perd dans le "bruit" stochastique.
  - **Inefficacité** : exploration large mais pas toujours utile, surtout quand les récompenses sont rares.

- La question n'est plus **d'explorer plus**, mais **d'explorer mieux**.
- Introduction de **Constrained Intrinsic Motivation (Zheng et al., 2024) :** une approche qui apprend à *réguler* sa propre curiosité.

CIM : Le Principe d'une Curiosité régulée

Récompense totale adaptative :

$$r_t = \alpha_t \times r_{int} + \beta_t \times r_{ext}$$

Les poids **α_t** (curiosité) et **β_t** (tâche) ne sont plus fixes. Ils sont ajustés dynamiquement en fonction :

- Du **progrès d'apprentissage** (ralentissement ou accélération).
- De la **stabilité** de la politique de l'agent.

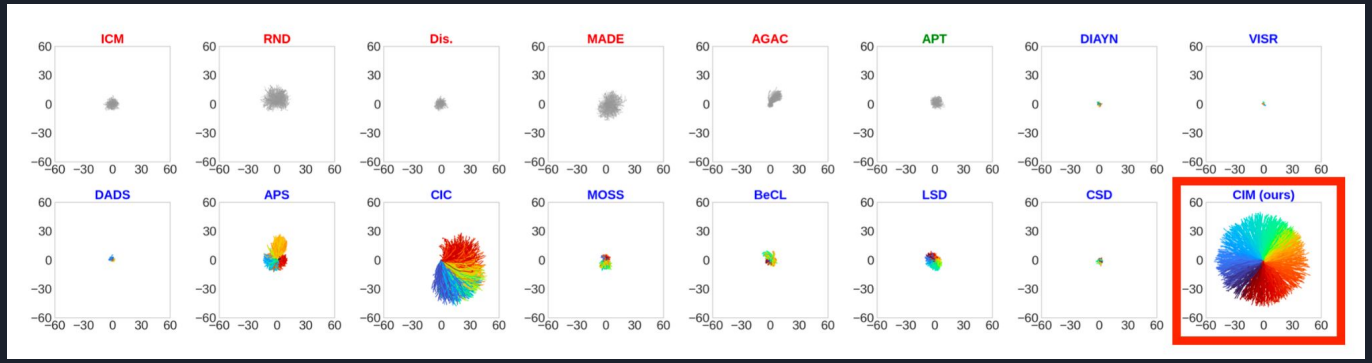
Objectif : Équilibrer dynamiquement exploration (utile) et exploitation (performance).
Contrainte d'alignement (L_a) : Force les compétences apprises à être à la fois dynamiques (provoquer du changement) et distinctes (ne pas être redondantes).

Résultats : Explorer Moins, Apprendre Mieux

Tests sur des benchmarks complexes (MuJoCo, AntMaze).
Supériorité par rapport à 15 autres méthodes de motivation intrinsèque.
Apprentissage 20x plus efficace en termes de nombre d'échantillons.
Compétences découvertes :

- Plus **dynamiques** (mouvements amples et utiles).
- Plus **diversifiées** (couvrent mieux l'espace des possibles).

Conclusion : La régulation de la curiosité permet une exploration **stratégique** plutôt qu'impulsive



Visualisation de compétences de locomotion continues en 2D sur l'environnement *Ant*.

Vers une Métacognition Artificielle

- **Inspiration humaine :** Nous ne sommes pas curieux en permanence.
- **Régulation de l'effort cognitif :** On se concentre quand on apprend, on se relâche quand on maîtrise.
- **Zone Proximale de Développement (Vygotsky) :** L'apprentissage est maximal quand la difficulté est "juste bonne". CIM recherche activement cette zone.
- **Métacognition :** La capacité à "penser sur ses propres pensées" et à évaluer son propre apprentissage.
- **CIM :** Une implémentation de la métacognition. L'agent évalue son propre progrès pour décider s'il doit explorer ou exploiter.

Défis Futurs : L'Horizon de l'Autonomie Cognitive

- CIM est une étape clé, mais la route est longue.
- **1. Curiosité Hiérarchique & Abstraite :**
  - Comment être curieux non seulement de "bouger un bras", mais d'un concept comme "construire un abri" ?
- **2. Mémoire à Long Terme & "Re-curiosité" :**
  - Comment un agent peut-il décider de redevenir curieux sur un sujet qu'il a délaissé ? (Lien avec la mémoire épisodique).
- **3. Alignement & Contrôle Éthique :**
  - Une IA qui régule ses propres objectifs d'apprentissage reste-t-elle alignée avec les buts humains ? C'est un enjeu de sécurité majeur.

Conclusion : De l'Exploration à l'Apprentissage Intelligent

- **Évolution de la recherche :**
  - **Phase 1 (ex: Oudeyer, 2007) :** Comment créer un moteur de curiosité ? (*Quoi explorer ?*)
  - **Phase 2 (Survey Aubret, 2023) :** Comment catégoriser les types de curiosité ? (Surprise, Nouveauté...)
  - **Phase 3 (ex: Zheng, 2024) :** Comment réguler intelligemment la curiosité ? (*Quand explorer ?*)
- La motivation intrinsèque n'est pas qu'une solution à l'exploration, c'est une brique fondamentale pour un **apprentissage autonome et efficace**.

Activité débat

1	2	3	4
Choix d'un sujet	**Choix de son camp** Pour ou Contre	**5 min Délibération intra-groupe** Trouver des arguments	**Débat**

Sujets

- Une IA curieuse, est-ce dangereux ?
- On imagine des IA d'assistance personnelle optimales comme apprenant par elle-même comment s'aligner selon nos désirs et anticiper nos besoins. Cela implique de les dôter d'une curiosité autonome, ce qui les rend plus imprévisibles et moins contrôlables. Doit-on sacrifier notre sécurité pour une meilleure utilité ?