

## Soutenance de thèse

**Paul TEMPLIER** soutiendra sa thèse de doctorat, préparée au sein de l'équipe d'accueil doctoral ISAE-ONERA MOIS et intitulée «*Synergies dans la recherche évolutionnaire de politiques connexionnistes*»

**Le 22 avril 2024 à 9h00, salle des thèses, ISAE-SUPAERO**

devant le jury composé de

M. Emmanuel RACHELSON	ISAE-SUPAERO	Directeur de thèse
M. Dennis WILSON	ISAE-SUPAERO	Co-directeur de thèse
M. Nikolaus HANSEN	INRIA	Rapporteur
M. Olivier SIGAUD	Sorbonne Université	Rapporteur
Mme Gabriela OCHOA	University of Stirling	
M. Daniel DELAHAYE	ENAC	

**Résumé :** Former un agent pour des tâches complexes comme conduire une voiture, maîtriser un jeu vidéo ou contrôler le plasma dans un réacteur à fusion nucléaire pose un défi important en l'absence d'un guide expert. Dans de tels scénarios, imiter le processus d'évolution du monde naturel, qui a affiné les capacités de résolution de problèmes des cerveaux animaux, offre une approche prometteuse. La Recherche Évolutionnaire de Politiques Connexionnistes (REPC) s'inspire de ce concept. Elle crée une population diversifiée d'agents, ou "politiques", représentés par des réseaux de neurones artificiels, permettant au système d'évoluer en sélectionnant et en mutant sélectivement les individus performants. Cette thèse explore les composants clés de la REPC et leurs interactions, en se concentrant sur les Stratégies Evolutionnaires (ES), une méthode d'optimisation stochastique basée sur une distribution. Nous présentons trois axes de recherche pour tirer parti des synergies entre ES et recherche de politiques dans le développement de nouvelles méthodes d'apprentissage. Le premier problème abordé est lié à l'incertitude dans l'évaluation des solutions, où le score d'une politique peut être affecté par du bruit ou un environnement stochastique. Pour propager l'incertitude depuis le score jusqu'à la mise à jour de la distribution de l'ES, nous proposons une méthode d'estimation de la dispersion par échantillonnage Monte-Carlo (DEMONS) qui identifie la solution ayant le plus grand impact et nécessitant plus d'échantillons pour être évaluée avec précision. Ce nouveau schéma d'évaluation tire parti de la structure de l'étape de mise à jour ES pour mieux identifier les solutions sensibles qui nécessitent plus d'attention. Notre second axe de recherche concerne l'exploitation de la structure des réseaux de neurones artificiels pour améliorer leur représentation en tant que génomes pour l'optimisation. Nous introduisons l'Encodage Géométrique pour la Neuroévolution (GENE) qui encode l'information dans les neurones, réduisant leur dimension pour rendre l'optimisation moins coûteuse tout en façonnant l'espace de recherche pour faire émerger naturellement de meilleures architectures. En utilisant une approche de méta-évolution, nous optimisons ensuite cet encodage pour obtenir une représentation des réseaux qui améliore les performances tout en généralisant sur de nouveaux problèmes. Enfin, nous explorons deux domaines adjacents, l'Apprentissage par Renforcement (RL) et la Qualité-Diversité (QD), qui utilisent des informations supplémentaires provenant des évaluations pour guider la recherche. Les transitions d'états rencontrées par les agents peuvent être utilisées par du RL pour entraîner une politique, l'acteur, qui est ensuite injecté dans la population. Cependant, nous pouvons montrer que la différence entre les objectifs de l'ES et du RL font que les génomes de l'acteur et de la population divergent, ce qui mène à une rupture de l'optimisation de l'ES. Nous introduisons la Régularisation de la Dérive Génétique, qui contraint l'acteur RL à rester génétiquement proche de la distribution ES, améliorant ainsi leurs performances QD définit des descripteurs de comportement pour chaque solution, qui y

sont utilisés pour créer un répertoire de politiques diverses. Nous montrons que ces descripteurs peuvent également être inclus dans le processus ES pour aider à trouver la meilleure solution. Nous introduisons le cadre de la Qualité avec Juste Assez de Diversité (JEDi), qui apprend le lien entre comportement et scores pour se concentrer sur les solutions prometteuses. JEDi utilise les comportements pour introduire plus de diversité dans la recherche, résolvant efficacement les tâches d'exploration difficiles, et pour trouver des étapes intermédiaires qui améliorent la recherche.

**Mots clés :** apprentissage automatique, machine learning, stratégies évolutionnaires, réseaux de neurones

**Summary:** Artificial Neural Networks (ANN) have fundamentally redefined the way data is analyzed and integrated with the new field of deep learning. Despite these advances, ANNs still have many problems that are not seen in biological neural networks, such as catastrophic forgetfulness. The idea of this project is to apply the learning principles observed in biological neurons to artificial neural networks, including synaptic plasticity and structural learning. This will allow neural networks to adapt their structure and learning to specific tasks and to learn continuously without forgetting previous knowledge. The objective of this thesis is to improve reinforcement learning methods. This fundamental work on ANN learning has advantages in all its fields of application.

**Keywords:** machine learning, evolution strategies, neural network