**Electricité de France R&D** Octobre 2019

Département ICAME

7 boulevard Gaspard Monge

92120 Palaiseau

*Proposition de stage 2019 :*

Interprétabilité de méthodes d’apprentissage par renforcement

*Descriptif :*

# Contexte général

 Ces dernières années les méthodes d’apprentissage par renforcement associées à des réseaux de neurones ont connu des succès importants. Dans le domaine des jeux ATARI, les équipes de Google DeepMind sont parvenues à atteindre des performances supérieures à celles de l’humain, en utilisant un algorithme de Deep-Q-Learning [1]. Plus récemment, ces mêmes équipes ont battu les meilleurs joueurs humains de GO avec un algorithme de MCTS [2].

Il s’agit dans les deux cas de créer un agent capable d’effectuer une série d’actions sur un environnement afin de maximiser un objectif. Cet agent possède une politique qui définit les actions à prendre en fonction de l’état de l’environnement, et une valeur qui estime le gain attendu à partir de cet état ([3]). La politique et la valeur sont représentées par un réseau de neurones, et ajustées lors d’une période d’apprentissage au cours de laquelle l’agent réalise une série d’expériences du type essai-erreur. Des bibliothèques open-source (keras, tensorflow) sont désormais disponibles et permettent d’implémenter facilement ces algorithmes.

Ce formalisme peut être exploité pour de multiples problématiques industrielles, notamment dans le cadre de problèmes d’optimisation, où l’optimum peut être atteint par actions d’un agent sur un système, ou pour des problèmes de recherche d’une stratégie optimale pour des systèmes personnalisables.

**Problème industriel**

Les départements ICAME, PRISME et PERICLES s’intéressent actuellement à l’utilisation de méthodes de renforcement pour traiter différents problèmes.

Le premier cas d’usage envisage concerne le rechargement des cœurs de réacteurs du parc nucléaire français : soumis à plusieurs jeux de contraintes physiques, la construction d’un plan de chargement s’apparente à un problème d’optimisation, où la recherche d’un optimum peut se faire par permutation successives des assemblages présents dans le cœur. Ce processus correspondant au formalisme de l’apprentissage par renforcement, des algorithmes basés sur du deep-q-learning et du Monte-Carlo Tree Search ont été testés en 2017 pour aborder cette question.

Pour l’année 2018 et dans le cadre de travaux sur les smarts grids et la gestion des réseaux de distribution d’énergie, il est également envisagé de s’intéresser à des solutions basées sur du renforcement pour le développement d’un agent s’adaptant au profil des consommateurs présents sur un réseau de distribution [4].

# Objectif du stage

 Dans le cadre du développement des méthodes précédentes, plusieurs modèles et architectures ont été testées pour l’algorithme de renforcement. Le choix des méthodes et des paramètres est actuellement principalement empirique, alors que leur impact apparaît crucial pour les performances des algorithmes. La mise en place d’un modèle optimal par rapport aux contraintes du problème posé et aux objectifs de résultats nécessite une évaluation plus rigoureuse des solutions testées et de leur paramétrage, afin d’analyser l’influence de ces choix sur le processus d’apprentissage.

 Dans ce cadre, l’objectif du stage sera donc de mettre en place des indicateurs de performances des modèles et d’évaluer l’influence des différents choix pour l’apprentissage du réseau de neurones ou pour l’expérimentation des agents sur ces indicateurs.

Le travail nécessitera notamment de :

* Comprendre les algorithmes utilisés et l’influence de leur paramétrage
* Mettre en place des indicateurs permettant d’évaluer la qualité des résultats obtenus par les différents modèles
* Développer des méthodes de visualisation permettant d’explorer le processus d’apprentissage

**Le contenu du stage pourra être adapté en fonction de l’avancée des travaux en cours et des compétences du stagiaire.**

*Références :*

[1] V. Minh (Google Deep Mind) et al., Human-level control through deep reinforcement learning, Nature, vol 518, pp 529-541, Février 2015

[2] D.Silver (Google Deep Mind) et al., Mastering the game of Go without human knowledge, Nature vol 550, pp 354-359, Octobre 2017

[3] R.Sutton et al, Reinforcement Learning, An Introduction, The MIT Press, 2017

[4] Mocanu, Elena, et al. "Unsupervised energy prediction in a smart grid context using reinforcement cross-building transfer learning." *Energy and Buildings* 116 (2016): 646-655.

*Conditions matérielles :*

Le stagiaire sera encadré par un ingénieur du groupe E72 **« Statistiques et Outils d’Aide à la Décision»**. L’encadrement du stagiaire pourra également s’appuyer sur des compétences disponibles dans d’autres départements, notamment PRISME et PERICLES.

Lieu du stage : EDF-R&D, 7 boulevard Gaspard Monge, 92120 Palaiseau

Durée : **6 mois**

Sujets abordés par le stage: **Intelligence artificielle, Physique des réacteurs nucléaires, Utilisation de langages scientifiques (Python, Keras, TensorFlow), Système d’exploitation Unix/Linux.**

Compétences requises :

* Langages de programmation : Python, R
* Librairies : Tensorflow, Keras, Pytorch
* OS : Linux, Windows
* Modèles :
* Réseaux de neurones : réseaux convolutifs et récurrents, auto-encodeurs, GAN
* Apprentissage par renforcement : Q\_learning, MCTS, policy gradient

*Renseignements complémentaires :*

**Pour plus d’information contacter :**

**Lou CHARAUDEAU**

**Tél. : +33.(0)1. *78 19 31 43***

**E-mail : lou.charaudeau@edf.fr**