Apprentissage par renforcement pour le calcul d'équilibres de Nash dans les jeux stochastiques – Application au calcul/recalcul d'objectifs de haut niveau dans les jeux de conservation stochastiques
Stage de Master 2

Encadrants : Jean Jane Kiam (UniBW), Caroline Chanel (ISAE-SUPAERO),
Régis Sabbadin et Meritxell Vinyals (Inrae)

Laboratoire d'accueil, encadrement

Un sujet de stage de Master 2 et un sujet de thèse en Intelligence Artificielle sont proposés dans le cadre du projet PRCI ANR-DFG *Coordinating Heterogeneous Interacting Planning Agents Using Game Theory* (**CHIP-GT**), dont les partenaires sont l'Inrae (Unité MIAT, Centre Occitanie-Toulouse), l'ISAE-SUPAERO (également localisée à Toulouse) et l'University of the Bundeswehr (Institute of Flight Systems), à Munich.

Le stage de M2 se déroulera au printemps 2023, dans l'Unité MIAT du Centre Inrae de Toulouse. La thèse se déroulera dans la même Unité et débutera en Septembre ou Octobre 2023. La ou le stagiaire de M2 pourra, selon son souhait, poser sa candidature pour la thèse.

Pour toute candidature, contacter meritxell.vinyals@inrae.fr ou regis.sabbadin@inrae.fr.

Sujet de stage

Les jeux stochastiques [4, 1] sont un cadre pour l'analyse des interactions entre agents non coopératifs évoluant dans un environnement dynamique, incertain et partiellement observé. Ils généralisent à la fois le cadre des *Processus Décisionnels Markoviens (PDM)* [5], dédié à la modélisation et la résolution de problèmes de planification dans l'incertain mono-agents et le cadre des *jeux non-coopératifs* [3], permettant de modéliser les interactions entre agents et de calculer des stratégies d'équilibre (équilibres de Nash) pour des problèmes de décision non-séquentielle.

Le calcul des stratégies d'équilibres dans un jeu stochastique est un problème difficile, en particulier lorsque le modèle du jeu est inconnu et seulement accessible via des interactions entre les joueurs et avec l'environnement, qui peut être réel ou simulé. Dans le contexte mono-agent, le domaine de *l'apprentissage par renforcement* [7] propose de nombreux résultats et outils de calcul de stratégies optimales pour des PDM dont le modèle est inconnu et seulement accessible par simulation. Dans le contexte des jeux stochastiques, la littérature est bien moins riche en outils de type apprentissage par renforcement pour le calcul de stratégies d'équilibre [2, 6].

Travail attendu

Une des tâches du projet ANR **CHIP-GT** consistera à explorer le cadre de l'apprentissage par renforcement dans les jeux stochastiques et à proposer et étudier de nouveaux algorithmes de calcul de stratégies d'équilibre. Cette tâche sera abordée, entre autres, par un doctorant qui sera recruté à l'automne 2023.

Afin de préparer ce travail le stagiaire de Master 2 recruté sera chargé de concevoir un environnement de simulation dédié aux jeux stochastiques en général (également appelés Markov games), instancié également dans le sous-domaine des jeux de conservation stochastiques, définis par les encadrants. Cet environnement sera construit en Python, en utilisant l'API Gym de la société OpenAI. Cette API est dédiée à la construction d'environnements de simulation pour le développement et l'évaluation d'algorithmes d'apprentissage

par renforcement. Bien que *Gym* soit dédié initialement à l'apprentissage par renforcement mono-agent, il permet également de créer des environnements dédiés aux problèmes d'apprentissage par renforcement multi-agents [1].

Le stagiaire devra tout d'abord se familiariser avec le modèle des jeux stochastiques et des processus décisionnels de Markov (sans s'attacher aux algorithmes de calcul de stratégies d'équilibre/optimales). Il devra également se familiariser avec l'API Gym, par exemple en recodant un environnement de simulation de stratégie dans un processus décisionnel Markovien simple. Ensuite, le stagiaire proposera un environnement de simulation de stratégies mixtes dans un jeu stochastique, complètement ou partiellement observé. Enfin, la personne recrutée participera à l'implémentation d'un modèle de jeu de conservation stochastique dans l'API Gym (ce modèle étant en cours d'élaboration, le stagiaire aura une certaine latitude pour le modifier).

Les environnements Gym développés seront testés avec différents agents simples, codés par l'étudiant : Stratégies jointes mixtes arbitraires, stratégies aléatoires... Si l'étudiant avance suffisamment vite, il pourra implémenter une stratégie classique d'apprentissage dans les jeux de type *fictitious play* [2].

Compétences requises

Ce stage requiert une certaine aisance avec des concepts mathématiques tels que les processus Markoviens, qui sont au coeur du cadre des Jeux stochastiques. Par ailleurs, une familiarité et une expérience avec le langage Python et la programmation objet est essentielle. Le code développé pendant le stage aura vocation à être réutilisé, amélioré, etc.

Des compétences en développement collaboratif (utilisation de GIT, tests unitaires, documentation...) seraient un plus, même si elles pourront être acquises pendant le stage.

Cadre de travail, rémunération

Le stage, d'une durée de 4 à 6 mois, se déroulera dans l'Unité MIAT de l'Inrae, à Toulouse. La durée hebdomadaire de travail est de 35h. Le stage donnera lieu à une indemnité d'environ 540 Euros par mois. Le Centre Inrae est facilement accessible par les transports en commun (ou en vélo), dispose d'un restaurant d'entreprise, et une association propose de nombreuses activités sportives et autres.

Références

[1] Jerzy Filar et Koos Vrieze. *Competitive Markov Decision Processes*. Springer-Verlag, 1996.

[2] Drew Fudenberg et David K. Levine. *The Theory of Learning in Games*. Cambridge, MA : MIT Press, 1998.

[3] John F Nash. « Equilibrium points in n-person games ». In : *Proceedings of the national academy of sciences* 36.1 (1950), p. 48-49.

[4] Abraham Neyman et Sylvain Sorin. *Stochastic Games and Applications*. Kluwer Academic Press, 2003.

[5] Martin L Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley et Sons, 1994.

[6] Karl Tuyls et Gerhard Weiss. « Multiagent Learning: Basics, Challenges and Prospects ». In : *AI Magazine* 33.3 (2012), p. 41-52.

[7] Kaiqing Zhang, Zhuoran Yang et Tamer Başar. « Vamvoudakis K.G., Wan Y., Lewis F.L., Cansever D. (eds) Handbook of Reinforcement Learning and Control ». In : t. 325. Studies in Systems, Decision and Control. Springer, 2021. Chap. Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms.

1. Voir https ://pythonawesome.com/a-simple-openai-gym-environment-for-single-and-multi-agent-reinforcement-learning/, par exemple.