

INVITATION

Apprentissage par Renforcement en Environnement Non-Stationnaire

Soutenance de thèse – Erwan Lecarpentier
Le 6 juillet 2020 à 10h

En raison de la crise sanitaire COVID-19, aucun public ne pourra assister en présentiel. La soutenance pourra être suivie en visio-conférence à partir de 10h via

<https://zoom.us/j/98739992593>

Devant le jury composé de :

Olivier Buffet,	chargé de recherche à l'INRIA, Nancy	Rapporteur
Tristan Cazenave,	professeur des universités à l'université Paris-Dauphine	Examineur
Aurélien Garivier,	professeur des universités à l'ENS Lyon	Examineur
Guillaume Infantes,	chercheur à Jolibrain, Toulouse	Directeur de thèse
Emilie Kaufmann,	chargée de recherche à l'INRIA, Lille	Examinatrice
Emmanuel Rachelson,	enseignant chercheur à l'ISAE-SUPAERO, Toulouse	Directeur de thèse
Régis Sabbadin,	directeur de recherche à l'INRA, Toulouse	Examineur
Bruno Zanuttini,	professeur des universités à l'université de Caen	Rapporteur

Invité :

Charles Lesire-Cabaniols, chercheur à l'ONERA, Toulouse Conseiller

Résumé :

Comment un agent doit-il agir étant donné l'incertitude qu'il a sur l'évolution de son environnement ? Dans cette thèse, nous adoptons la perspective de l'apprentissage par renforcement pour fournir une réponse à cette question. Le problème est vu sous trois aspects différents. Premièrement, nous étudions le compromis planification vs. re-planification des algorithmes de recherche arborescente dans les Processus Décisionnels Markoviens stationnaires. Nous proposons une méthode pour réduire la complexité de calcul d'un tel algorithme, tout en conservant des garanties théoriques sur la performance. Deuxièmement, nous étudions le cas des environnements évoluant graduellement au cours du temps. Cette hypothèse est formulée dans un cadre mathématique appelé Processus de Décision Markoviens Non-Stationnaires Lipschitziens. Dans ce cadre, nous proposons un algorithme de planification robuste aux évolutions possibles, dont nous montrons qu'il converge vers la politique minmax. Troisièmement, nous considérons le cas de l'évolution temporelle abrupte dans le cadre du "lifelong learning" (apprentissage tout au long de la vie). Nous proposons une méthode de transfert non-négatif basée sur l'étude théorique de la continuité de Lipschitz de la Q-fonction optimale par rapport à l'espace des tâches. L'approche permet d'accélérer l'apprentissage dans de nouvelles tâches. Dans l'ensemble, cette dissertation propose des réponses à la question de la résolution des Processus de Décision Markoviens Non-Stationnaires dans trois cadres d'hypothèses.

Mots-clés :

planification ; apprentissage par renforcement ; Processus Décisionnel de Markov ; Processus Décisionnel de Markov Non-Stationnaire ; apprentissage tout au long de la vie.