Soutenance de thèse

Giorgio ANGELOTTI soutiendra sa thèse de doctorat, préparée au sein de l'équipe d'accueil doctoral ISAE-ONERA DECISIO et intitulée «*Contributions à l'apprentissage par renforcement hors ligne avec prise en compte du risque : étude sur l'augmentation des données, sur la sélection des politiques et sur l'explicabilité*»

Le 12 juin 2023 à 14h00, salle des thèses ISAE-SUPAERO

devant le jury composé de

Mme Caroline PONZONI CARVALHO CHANEL	Professeure Associée ISAE-SUPAERO	Directrice de thèse
M. Marek PETRIK	Associate Professor University of New Hampshire	Rapporteur
M. Bruno ZANUTTINI	Professeur Université de Caen Normandie	Rapporteur
M. Emmanuel RACHELSON	Professeur ISAE-SUPAERO	Examinateur
M. Vincent THOMAS	Maître de conférences Université de Lorraine	Examinateur
M. Régis SABBADIN	Directeur de recherche MIAT INRAE Toulouse	Examinateur
M. Nicolas DROUGARD	Professeur Associé ISAE-SUPAERO	Co-encadrant de thèse
M. Rémi MUNOS	Directeur de recherche INRIA Lille	Examinateur

Résumé : Dans le domaine de l'apprentissage par renforcement hors ligne, l'objectif est d'apprendre une politique de décision hors ligne, c'est-à-dire sur la base d'un lot d'expériences collectées précédemment et sans interaction supplémentaire, de préférence d'une manière efficace en termes de données et sensible au risque. Cette thèse présente plusieurs techniques pour atteindre cet objectif, en mettant l'accent sur les méthodes basées sur des modèles : des paradigmes qui infèrent d'abord un modèle comportemental pour le problème de prise de décision séquentielle et le résolvent ensuite en prenant en compte l'incertitude de l'estimation du modèle. Les contributions présentées comprennent une méthode pour augmenter un ensemble de données d'échantillons en détectant les symétries dans la dynamique du système, une méthode pour effectuer une sélection de politique sensible au risque hors ligne appelée Exploitation vs Caution (EvC) en recourant au cadre du processus de décision de Markov bayésien, et un paradigme pour l'explicabilité dans les systèmes coopératifs multi-agents en utilisant l'analyse de Myerson. De plus, nous discutons des perspectives d'application de l'approche EvC pour obtenir une politique de contrôle d'interaction adaptative dans un scénario homme-robot. En effet, en prenant les précautions nécessaires, nous avons adapté l'algorithme EvC pour la sélection de politiques sensibles au risque afin de l'appliquer au ISAE Robot Firefighter Game, qui implique l'optimisation de politiques adaptatives pour contrôler l'interaction entre un robot pompier et un pompier humain dans un scénario de preuve de concept. Dans l'ensemble, les contributions de cette thèse démontrent le potentiel des techniques présentées pour améliorer de manière significative la performance des algorithmes d'apprentissage par renforcement hors ligne et pour être appliquées dans une variété de contextes du monde réel, y compris l'interaction homme-robot.

Mots-clés : apprentissage par renforcement hors ligne, sensible au risque, interaction homme-robot, augmentation des données, explicabilité, sélection des politiques

Summary : Recent advances in Machine Learning and Robotics are paving the way to an increasing adoption of automated vehicles in everyday life. Nevertheless, leaving full decision-making autonomy to an automated agent is still a gamble given the current performances of state-of-the-art planners and Reinforcement Learning algorithms. Sometimes both ethical and legal reasons involve the presence of a human operator in the decision-making process. The human operator is traditionally seen as a fool-proof agent to whom has been granted the authority to take complete control of the system at any moment. In this thesis we rather focus on the Mixed-Initiative Team composed of a human operator and automated agent. In such a setting the human and the automated agent stand on equal footing and cooperate to accomplish a mission. More specifically we address the problem of optimizing the interaction between a firefighter robot and a fireman in a proof-of-concept serious game previously developed at ISAE Supaero. We devise a methodology to implement an additional automated control system that drives the interaction, allocating the initiative and the tasks between the human and the robot with the aim of improving the mission performance. Since there is not something as a general behavioural model of a human being, planning with a human in the loop is a considerable challenge. In a data-driven fashion we learn a Mixed-Interaction model shaped using the mathematical framework of Partially Observable Markov Decision Processes (POMDPs) to include the inner condition of the human operator as a hidden state only partially observable through measurements of physiological features and behavioural traits. The model learning phase runs completely offline and exploits a pre-collected data set of trajectories: sequences of observations of both the system and the human operator and the synchronized actions applied by the control system. We resort to the paradigm of Bayesian Risk-Sensitive Optimisation to cope with possible diversity and scarcity of the data set that would hinder the inference of a realistic one-step model and result, after the subsequent optimisation, into risky and dangerous control policies. Finally, we want to show through laboratory experiments with simulated fires, a real robot and multiple human volunteers that not only the proposed Mixed-Initiative control system provide more robust and performing policies with respect of a full manual (human driven) or a full automatic (robot driven) setting, but also that real time measurements of physiological and behavioural traits help the planning task while the Bayesian Risk-Sensitive optimisation improves the robustness of the approach.

Keywords : offline reinforcement learning, risk-sensitive, human-robot interaction, data augmentation, explainability, policy selection