Soutenance de thèse

Arthur CLAVIÈRE soutiendra sa thèse de doctorat, préparée au sein de l'équipe d'accueil doctoral ISAE-ONERA MOIS et intitulée « *Vérification de la sûreté des systèmes basés réseaux de neurones s'appuyant sur les méthodes formelles* »

Le 17 juillet 2023 à 10h00 – salle des thèses – ISAE-SUPAERO

devant le jury composé de

Mme Claire PAGETTI	Directrice de recherche ONERA	Directrice de thèse
M. Éric GOUBAULT	Professeur École Polytechnique	Rapporteur
Mme Susanne GRAF	Directrice de recherche VERIMAG	Rapporteure
M. Joao MARQUES-SILVA	Directeur de recherche IRIT	
Mme Elisa FROMONT	Professeure Université de Rennes/IRISA	
M. Éric ASSELIN	Docteur Collins Aerospace	Co-directeur de thèse du monde socio-économique
M. Benedikt BOLLIG	Directeur de recherche LMF	

Résumé : Contexte : la thèse a porté sur l'étude et la vérification de la sûreté de fonctionnement de systèmes contrôlés par réseaux de neurones. Ce type de système combine un système physique et un contrôleur basé réseaux de neurones. L'utilisation de ce type de contrôleur peut avoir différents intérêts : (1) approximer un autre contrôleur, déjà existant, mais en demandant moins de ressources de calcul (sachant qu'un contrôleur dispose souvent de ressources limitées) ou (2) reproduire le comportement d'un humain (ce qui peut s'avérer intéressant pour les systèmes autonomes). Dans le cas où le système contrôlé par réseaux de neurones est critique i.e., une défaillance de ce système peut entraîner des conséquences graves, alors il est important de vérifier sa sûreté de fonctionnement i.e., montrer qu'il n'atteint pas d'états non sûrs. Hypothèses : Dans la lignée des cas d'études présents dans la littérature, nous avons restreint notre étude à une forme particulière de contrôleur basé réseaux de neurones : un classificateur basé sur un ou plusieurs réseaux de neurones. Par ailleurs, nous avons considérés uniquement des réseaux de neurones feed-forward fully connected avec des fonctions d'activation ReLU. Exemple : un exemple typique de ce type de système est l'ACAS Xu qui comprend un contrôleur basé réseaux de neurones dont le rôle est d'éviter une collision entre deux avions. Pour ce système, le contrôleur basé réseaux de neurones approxime un autre contrôleur mais avec une empreinte mémoire très réduite. C'est un système critique car une défaillance peut entraîner une collision. Contributions : Afin de démontrer la sûreté de fonctionnement du système d'intérêt, nous avons d'abord considéré le cas où le contrôleur basé réseaux de neurones approxime un autre contrôleur. Pour ce cas de figure, nous avons développé une méthode et un outil afin de comparer les deux contrôleurs et montrer que le contrôleur basé réseaux de neurones est une approximation correcte du contrôleur cible. Notre méthode et notre outil ont été appliqués à l'ACAS Xu. Pour ce système en particulier, nous avons montré que nos travaux pouvaient servir au développement d'un contrôleur hybride, combinant le contrôleur basé réseaux de neurones et le contrôleur original, offrant à la fois une empreinte mémoire réduite et un comportement sûr. Pour notre seconde contribution, nous nous sommes concentrés sur la vérification du système contrôlé par réseaux de neurones en entier, pas seulement le contrôleur. Cette seconde contribution comporte deux aspects : (i) le développement d'un modèle basé automate hybride du système étudié et (ii) un outil, appelé SAMBA,

qui permet l'analyse de cet automate hybride. Nos travaux ont été appliqués à l'ACAS Xu et nous avons aussi mené des comparaisons poussées avec d'autres outils de l'état de l'art. Ces comparaisons ont été menées sur la base de trois cas d'usages, parmi lesquels l'ACAS Xu, présentant différents types de dynamique, linéaire ou non linéaire, et impliquant des réseaux de neurones de différentes tailles. Les outils ont été comparés notamment sur leur capacité à résoudre un large éventail de problèmes de vérification. Nous avons aussi pu tirer plusieurs leçons de ces expériences et une méthodologie pour choisir le meilleur outil pour un problème de vérification donné.

Mots clés : Réseaux de neurones, Sûreté, Méthodes formelles

Summary: Machine learning based techniques pave the way for a new generation of embedded applications aboard aircraft, UAVs, helicopters, cars and others, allowing to imagine the future of decision making and piloting, autonomous systems and other unsuspected technologies. However, these applications face a major obstacle: providing guarantees about their safety. Traditional assurance activities are not compatible with the intrinsic nature of machine learning based algorithms, due to their unpredictable behaviour for example. In this context, the thesis focuses on three main objectives, based on an aeronautical case study: defining the notion of safety for a given application, in line with the case study, proposing a tooled approach covering methodological and technical aspects for demonstrating the proposed concept of safety, and finally propose an integration of the approach into a certification process.

Keywords: Neural networks, Safety, Formal methods