Devising certifiable and explainable algorithms for verification and planning in cyber-physical systems

Résultats en bref

Expliquer le résultat d'un algorithme

Avant que les algorithmes ne puissent être utilisés dans des opérations critiques comme celles qui permettent de contrôler le trafic aérien, les utilisateurs humains doivent pouvoir leur faire confiance. Des explications, baptisées «certificats», sur les raisons pour lesquelles les résultats d'un algorithme sont corrects, pourraient servir de base pour instaurer cette confiance.

ÉCONOMIE NUMÉRIQUE

© Asset data, Shutterstock

Les algorithmes correspondent à un ensemble de règles ou d'instructions qui indiquent à un ordinateur ce qu'il doit faire. Bien que ces processus jouent un rôle essentiel en aidant les développeurs de logiciels à créer des programmes efficaces et exempts d'erreurs, ils ont aussi leurs limites, en particulier leur fréquente incapacité à apporter une preuve tangible de leurs résultats.

Selon Shaull Almagor, professeur adjoint d'informatique au Technion, cette carence en matière de certification représente un véritable défi, notamment pour la conception des systèmes cyber-physiques de demain.

«Prenons le cas d'un algorithme de planification robotique, où l'on demande à un robot d'entreprendre une tâche, consistant par exemple à trouver une clé et ouvrir une porte», explique-t-il. «Supposons maintenant que l'algorithme indique à

l'utilisateur qu'il n'existe aucune façon de guider le robot pour suivre un plan de ce type. Comment l'utilisateur peut-il savoir que cette réponse est correcte?»

Ou alors, qu'en est-il si l'algorithme trouve un plan, mais qu'il s'avère très compliqué. Sur quelle base l'utilisateur peut-il faire confiance à ce plan? «Ce problème est extrêmement important pour les situations où un superviseur humain doit approuver une opération critique pour la sécurité, par exemple dans les cas du contrôle du trafic aérien ou d'une usine industrielle», explique Shaull Almagor. «Dans ce type de situation, il faudrait idéalement que l'algorithme soit capable d'expliquer clairement pourquoi ses résultats sont corrects.»

Avec le soutien du projet ALGOCERT, financé par l'UE, Shaull Almagor a entrepris de concevoir des explications de ce type, qu'il appelle certificats.

Adapter les certificats à des contextes spécifiques

Comme l'explique Shaull Almagor, pour être pertinents, ces certificats doivent être adaptés à des contextes spécifiques. «Afin de définir les notions pertinentes des certificats, nous devons nous appuyer sur les caractéristiques spécifiques de la tâche de planification», ajoute-t-il. «C'est une condition préalable indispensable pour tenter de concevoir des algorithmes certifiables.»

Prenons par exemple le problème de l'atteignabilité pour les systèmes dynamiques linéaires représentant l'évolution de vecteurs au cours du temps. Bien que ces systèmes utilisent des règles mathématiques simples, ils peuvent donner lieu à des comportements très complexes.

«Si leur évolution leur permet d'atteindre un objectif, la certification est simple: il suffit d'indiquer le temps requis pour y parvenir», explique Shaull Almagor. «Mais si l'objectif n'est pas atteint, il n'est pas évident de proposer un certificat.»

Partant de ce constat, le projet a proposé une notion de certificats pour les scénarios où l'objectif ciblé n'est pas atteint. «Nous synthétisons un ensemble qui contient toute l'évolution du système mais qui ne recoupe pas la cible», ajoute Shaull Almagor.

Un autre problème abordé par le projet est celui de la recherche de chemin multi-agent (MAPF pour «pathfinding multi-agent»). Selon Shaull Almagor, la MAPF représente un problème fondamental pour la planification robotique, notamment pour des applications robotiques comme les entrepôts et les véhicules automatisés.

«Ces applications nécessitent que plusieurs agents soient capables de suivre simultanément des chemins établis sans entrer en collision les uns avec les autres», note Shaull Almagor.

Le projet a proposé une façon schématique d'expliquer à un utilisateur humain la justesse d'un plan MAPF. «Nous convainquons l'utilisateur que le plan ne présente aucun risque de collision en le décomposant sous la forme d'une séquence d'images, les chemins des agents étant disjoints dans chaque image», explique Shaull Almagor. «C'est vraiment une façon intuitive d'expliquer la non-collision.»

Un palier important

Ce ne sont là que quelques exemples parmi les problèmes abordés au cours du projet ALGOCERT, soutenu par le programme Actions Marie Skłodowska-Curie. Beaucoup d'entre eux ont fait l'objet d'explications détaillées dans divers articles et présentations.

«La notion d'explicabilité a suscité beaucoup d'attention au cours des dernières années», conclut Shaull Almagor. «Je n'ai pas la prétention d'affirmer que nos méthodes constituent le Saint Graal de la certification des algorithmes, mais je pense qu'elles représentent un palier important en ce sens.»

Mots-clés

ALGOCERT | algorithmes | explications | certificats | certification | logiciels | systèmes cyber-physiques | robots | opérations critiques pour la sécurité | systèmes dynamiques | mathématiques | recherche de chemin multi-agents

Informations projet

ALGOCERT

N° de convention de subvention: 837327

Site Web du projet

DOI
10.3030/837327

Projet clôturé

Date de signature de la CE
29 Avril 2019

Date de début
15 Juillet 2019

Date de fin
14 Juillet 2021

Financé au titre de
EXCELLENT SCIENCE - Marie Skłodowska-Curie Actions

Coût total
€ 185 464,32

Contribution de l'UE
€ 185 464,32

Coordonné par
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY
Israel

Dernière mise à jour: 10 Decembre 2021

Permalink: https://cordis.europa.eu/article/id/435398-explaining-an-algorithm-s-output/fr