HORIZON 2020

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

Ergebnisse in Kürze

Das Ergebnis eines Algorithmus erklären

Bevor Algorithmen in kritischen Bereichen wie der Flugsicherung eingesetzt werden können, müssen menschliche Nutzerinnen und Nutzer ihnen vertrauen können. Sogenannte Zertifikate, die erklären warum die Ergebnisse eines Algorithmus korrekt sind, könnten eine Basis für dieses Vertrauen schaffen.

DIGITALE WIRTSCHAFT

© Asset data, Shutterstock

Algorithmen bilden eine Reihe von Regeln oder Anweisungen, die einem Computer sagen, was er tun soll. Diese Prozesse spielen zwar eine wesentliche Rolle bei der Erstellung effizienter und fehlerfreier Programme, doch stoßen sie auch an ihre Grenzen. Besonders problematisch ist, dass sie oft keinen greifbaren Beweis für ihre Ergebnisse liefern können.

Laut Shaull Almagor, Assistenzprofessor für Informatik am Technion, stellt diese fehlende Zertifizierung ein echtes Hindernis dar, insbesondere bei der Entwicklung der cyber-physischen Systeme von morgen.

„Nehmen wir zum Beispiel einen Planungsalgorithmus für einen Roboter, mit dem dieser eine Aufgabe ausführen soll – z. B. einen Schlüssel suchen und eine Tür öffnen“, erklärt er. „Nehmen wir an, der Algorithmus teilt der nutzenden Person mit, dass es keinen solchen Plan für den Roboter gibt – wie kann sie wissen, ob diese

Antwort richtig ist?“

Oder wie geht man damit um, wenn der Algorithmus zwar einen Plan findet, dieser aber sehr kompliziert ist. Wie kann die Nutzerin oder der Nutzer diesem Plan vertrauen? „Dieses Dilemma kommt besonders zum Tragen in Situationen, in denen ein Mensch bei der Aufsicht einen sicherheitskritischen Vorgang genehmigen muss, wie z. B. bei der Flugsicherung oder in einem Industriewerk“, sagt Almagor. „In solchen Fällen wäre es ideal, wenn der Algorithmus eine einfache Erklärung dafür liefern könnte, warum sein Ergebnis korrekt ist.“

Mit Unterstützung des EU-finanzierten Projekts ALGOCERT machte sich Almagor an die Entwicklung solcher Erklärungen, die er als Zertifikate bezeichnet.

Anpassung von Zertifikaten an bestimmte Kontexte

Wie Almagor deutlich macht, müssen diese Zertifikate auf bestimmte Kontexte zugeschnitten sein, wenn sie aussagekräftig sein sollen. „Für sinnvolle Begriffsdefinitionen der Zertifikate müssen wir auf den spezifischen Eigenschaften der Planungsaufgabe aufbauen“, fügt er hinzu. „Erst dann können wir versuchen, nachweisbare Algorithmen zu entwickeln.“

Nehmen wir zum Beispiel das Problem der Erreichbarkeit für lineare dynamische Systeme, welche die zeitliche Entwicklung von Vektoren darstellen. Obwohl diese Systeme auf einfachen mathematischen Regeln beruhen, können sie zu sehr komplexen Verhaltensweisen führen.

„Wenn die Entwicklung ein Ziel erreicht, kann dies problemlos bescheinigt werden – es muss lediglich die Zeit angegeben werden, in der es erreicht wird“, stellt Almagor fest. „Wenn sie das Ziel aber nicht erreicht, ist unklar, wie ein Zertifikat ausgestellt werden kann.“

Auf dieser Grundlage schlug das Projektteam einen Zertifikatsbegriff für Szenarien vor, in denen das Ziel nicht erreicht wird. „Wir synthetisieren einen Satz, in dem die gesamte Entwicklung des Systems enthalten ist, der aber das Ziel nicht schneidet“, ergänzt Almagor.

Ein weiteres Problem, mit dem sich das Projekt befasst, liegt im Multi-Agent Path-Finding (MAPF). Almagor ist der Meinung, dass MAPF ein grundlegendes Problem bei der Planung von Robotern darstellt, einschließlich Roboteranwendungen wie z. B. automatisierte Lagerhäuser und Fahrzeuge.

„Diese Anwendungen erfordern, dass mehrere Agenten gleichzeitig den eingerichteten Pfaden folgen können, ohne ineinander zu stoßen“, merkt Almagor an.

In diesem Zusammenhang schlug das Projekt ein Verfahren vor, das einem nutzenden Menschen die Korrektheit eines MAPF-Plans näher bringt. „Wir überzeugen die Nutzerin oder den Nutzer davon, dass der Plan kein Kollisionsrisiko birgt, da er in eine Abfolge von Bildern zerlegt wird, bei der die Pfade der Agenten in jedem Bild nicht miteinander verbunden sind“, so Almagor. „Das ist wirklich eine intuitive Art, den ausbleibenden Zusammenstoß zu erklären.“

Ein wichtiges Sprungbrett

Damit sind nur einige der Probleme beschrieben, mit denen sich das durch die Marie-Skłodowska-Curie-Maßnahmen geförderte Projekt ALGOCERT befasst hat und von denen viele in verschiedenen Veröffentlichungen und Präsentationen ausführlich erläutert wurden.

„Der Begriff der Erklärbarkeit hat in den letzten Jahren viel Aufmerksamkeit erregt“, schließt Almagor. „Obwohl ich nicht behaupten kann, dass unsere Methoden der heilige Gral der Algorithmuszertifizierung sind, glaube ich doch, dass sie einen wichtigen Schritt auf dem Weg dorthin darstellen.“

Schlüsselbegriffe

ALGOCERT | Algorithmen | Erklärungen | Zertifikate | Zertifizierung | Software | cyber-physische Systeme | Roboter | sicherheitskritischer Vorgang | dynamische Systeme | mathematisch | Multi-Agent Path-Finding | MAPF

Projektinformationen

ALGOCERT

ID Finanzhilfevereinbarung: 837327

Projektwebsite

DOI
10.3030/837327

Projekt abgeschlossen

EK-Unterschriftsdatum

Finanziert unter
EXCELLENT SCIENCE - Marie Skłodowska-Curie Actions

Gesamtkosten
€ 185 464,32

EU-Beitrag
€ 185 464,32

Koordiniert durch
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY

29 April 2019

Israel

Startdatum	Enddatum
15 Juli 2019	14 Juli 2021

Letzte Aktualisierung: 10 Dezember 2021

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