Compositional Verification and Run-time Monitoring for Learning-Enabled Autonomous Systems

Corina Păsăreanu[1],
Ravi Mangal, Divya Gopinath

[1]Carnegie Mellon University and NASA Ames

Abstract

Providing safety guarantees for autonomous systems is difficult as these systems operate in complex environments that require the use of learning-enabled components, such as deep neural networks (DNNs) for visual perception. DNNs are hard to analyze due to their size (they can have thousands or millions of parameters), lack of formal specifications (DNNs are typically learnt from labeled data, in the absence of any formal requirements), and sensitivity to small changes in the environment. We present compositional techniques for the formal verification of safety properties of such autonomous systems. The main idea is to abstract the hard-to-analyze components of the autonomous system, such as DNN-based perception and environmental dynamics, with either probabilistic or worst-case abstractions. This makes the system amenable to formal analysis using off-the-shelf model checking tools, enabling the derivation of specifications for the behavior of the abstracted components such that system safety is guaranteed. The derived specifications are used as run-time monitors deployed on the DNN outputs. We illustrate the idea in a case study from the autonomous airplane domain.

AAAI 2024 Spring Symposium on User-Aligned Assessment of Adaptive AI Systems. Stanford University, Stanford, CA, USA.