Formal Verification of Continuous-Learning Cyber-Physical Systems

Thomas Flinkow

Supervised by Rosemary Monahan and Barak A. Pearlmutter

Abstract

Machine learning (ML) has recently achieved or even surpassed human capabilities in certain areas such as natural language processing and image classification, indicating great potential for integrating ML components into cyber-physical systems (CPSs) such as autonomous vehicles and aircraft.

In safety-critical domains, formal verification is used to ensure correctness and dependability. The past few years have seen the emergence of an extensive ecosystem of verification tools for neural networks (NNs), utilising a variety of approaches, such as constraint satisfiability and abstract interpretation. Yet, most tools are limited to an open-loop setting, checking only linear input-output constraints and (local) robustness, which fail to sufficiently capture the behaviour of a complex CPS where component-level specifications are frequently unavailable or entirely nonexistent.

Moreover, the majority of verification approaches assume trained models that cease learning post-deployment. A promising direction for verifying learning systems are so-called differentiable logics, which translate logical constraints into additional loss terms. Depending on the choice of domain and the specific logic, the expressivity of differentiable logics is limited and fails to fully capture the safety behaviour of a continuous-learning CPS. For instance, uncertainties inherent to the ML process may call for probabilistic specifications, while concepts such as resilience to stressors might require temporal and behavioural specifications.

Our research will investigate more expressive differentiable logics and assess their suitability for creating correct-by-construction machine-learnt models for use in CPSs. More generally, our research seeks to develop strategies and associated tooling for rigorously specifying continuous-learning CPSs and verifying them against rich and expressive specifications.