

# Specification Inference and Invariant Generation: A Machine Learning Perspective

Aditya V. Nori

Microsoft Research India

## Abstract

Computing good specification and invariants is key to effective and efficient program verification. In this talk, I will describe our experiences in using machine learning techniques (Bayesian inference, SVMs) for computing specifications and invariants useful for program verification. The first project Merlin uses Bayesian inference in order to automatically infer security specifications of programs. A novel feature of Merlin is that it can infer specifications even when the code under analysis gives rise to conflicting constraints, a situation that typically occurs when there are bugs. We have used Merlin to infer security specifications of 10 large business critical web applications. Furthermore, we show that these specifications can be used to detect new information flow security vulnerabilities in these applications.

In the second project Interpol, we show how interpolants can be viewed as classifiers in supervised machine learning. This view has several advantages: First, we are able to use off-the-shelf classification techniques, in particular support vector machines (SVMs), for interpolation. Second, we show that SVMs can find relevant predicates for a number of benchmarks. Since classification algorithms are predictive, the interpolants computed via classification are likely to be relevant predicates or invariants. Finally, the machine learning view also enables us to handle superficial non-linearities. Even if the underlying problem structure is linear, the symbolic constraints can give an impression that we are solving a non-linear problem. Since learning algorithms try to mine the underlying structure directly, we can discover the linear structure for such problems. We demonstrate the feasibility of Interpol via experiments over benchmarks from various papers on program verification.