CECS

CENTER FOR EMBEDDED & CYBER-PHYSICAL SYSTEMS
UNIVERSITY OF CALIFORNIA · IRVINE

CECS Seminar

"Safety Verification and Training for Learning-enabled Cyber-Physical Systems"

Jyotirmoy Vinay Deshmukh

Assistant Professor, Viterbi School of Engineering,
University of Southern California, Los Angeles

Tuesday, May 14, 2019
2:00 p.m. – 3:00 p.m.
Donald Bren Hall 4011

Abstract: With the increasing popularity of deep learning, there have been several efforts to use neural network based controllers in cyber-physical system applications. However, neural networks are equally well-known for their lack of interpretability, explainability and verifiability. This is especially an issue for safety-critical cyber-physical systems such as unmanned aerial vehicles or autonomous ground vehicles. How can we verify that a neural network based controller will always keep the system safe? We look at a new verification approach based on automatically synthesizing a barrier certificate for the system to prove that: starting from a given set of initial conditions, the system behavior can never reach an unsafe state. Barrier Certificates are essentially a generalization of inductive invariants for continuous dynamical systems, and we will show how we can use nonlinear SMT solvers to establish the barrier certificate conditions. A more intriguing challenge is whether we can actually train neural networks to obey safety constraints. We will look at a new way of reward shaping in reinforcement learning that could help achieve this goal.

Biography: Jyotirmoy V. Deshmukh (Jyo) is an assistant professor in the Department of Computer Science in the Viterbi School of Engineering at the University of Southern California in Los Angeles, USA. Before joining USC, Jyo worked as a Principal Research Engineer in Toyota Motors North America R&D. He got his Ph.D. degree from the University of Texas at Austin and was a post-doctoral fellow at the University of Pennsylvania. Jyo's research interest is in the broad area of formal methods. Currently, Jyo is interested in using logic-based methods for machine learning, and in techniques for the analysis, design, verification and synthesis of cyber-physical systems, especially those that use AI-based perception, control and planning algorithms.