Verifying Deep Neural Networks in Autonomous Cyber-Physical Systems: Dr. Taylor Johnson Colloquium Talk

Caleb Bowers

I. Background

Dr. Taylor Johnson is a professor of Computer Engineering, Computer Science, and Electrical Engineering in the School of Engineering at Vanderbilt University. He possesses broad research interests, but is primarily concerned with researching methods to enable the formal verification of cyber-physical systems. His worked has been recognized within academia and industry as a leading standard, and he has received funding from both government and industry programs.

II. Colloquium Overview

A. Cyber-Physical Systems and Deep Neural Networks

Dr. Johnson began his presentation with a discussion regarding the safety and maturity of self-driving cars (i.e., a cyber-physical system). He polled the audience to gauge the level of trust audience members place in the safety of self-driving cars. Unanimously, the audience believes that humans are better drivers, and Dr. Johnson was able to provide statistical backing for this position. Using this to segue into detailing the state-of-the-art in deep neural networks (DNN) and their applications (DNNs are used extensively in autonomous vehicles, so they provide an obvious area of verification research), he explained that the DNNs are used primarily for image classification and computer vision problem areas, which is, in part, why they are so heavily relied on in cyber-physical systems. Since DNNs provide a significant amount of the autonomy a cyber-physical system may possess, they must operate with some assurances and guarantees on the safety of the system, and in order to ensure these guarantees are met, the DNNs must be formally verified, which provides the basis for the majority of Dr. Johnson's presentation.

B. Deep Neural Network Verification

When verifying DNNs, safety properties are of primary concern, since these networks, when used in cyber-physical systems, can create courses of action that impact the safety of humans. Designers and engineers must be able to implement these DNNs with guarantees in place. In order to provide these guarantees on DNN safety, Dr. Johnson and his team have developed several methods and approaches to examine the properties that a given DNN could satisfy. The main focus of his research is to determine if the output space of a DNN inhabits the property space for a given system and a given property. A property for a DNN is stated in some logic formalism and using Logic Equivalence Checking (LEC) combined with a transformation (I believe) of the activation function of the DNN into what is called a Rectified Linear Unit (RELU) activation function, the DNN output space can be checked against the property space of the DNN as a whole. This section of the talk was a bit hard to follow, and I would love to look at the slides. I have actually emailed Dr. Johnson for a copy and I am awaiting his reply.

The fundamental result here is that for a given system a property can be defined for all time or for some given moment in time, and using this as the property space a simulation for the DNN and system can be conducted to determine if the output space of the DNN satisfies that property. By way of example, one could imagine an unmanned vehicle that needs to avoid an obstacle. The property could be: "Vehicle avoids obstacle," which is testably true or false. Running the system through a simulation of time will provide a look into whether or not this system will avoid the obstacle.

C. Brief Results

Dr. Johnson provided a brief overview of implementing his verification tools in Airborne Collision Avoidance System (ACAS) and in a project for Northrop Grumman using unmanned underwater vehicles (UUVs), both of which used DNNs for some motion directional action calculations. The Northrop Grumman test case demonstrated that for motion through space, Dr. Johnson's tools can produce the output space for a DNN reliant system

and how it reacts to an obstacle it needs to avoid. The ACAS examinations produced similar verifiability results for a smaller network used to inform platform collision avoidance decisions.

III. Conclusion

Dr. Johnson's talk was very interesting and I enjoyed being able to understand more naturally the formal methods and vernacular that he used to build up tools to help verify and understand DNNs. He laid the groundwork for his future work and the direction he would like to take. I asked him a question regarding what feedback information his methods produce in order to train the DNN better or give insight into the hidden layers of the network, and he expressed the intention that his work would take this direction with feedback information for DNN training being a principle goal for his work.